

# Un percorso: materiali per gli studenti

## 1 Motivazioni alla normale

### 1.1 I sondaggi

Per cercare di stabilire la posizione di una popolazione numerosa riguardo ad una determinata questione, sono molte le istituzioni che effettuano sondaggi<sup>1</sup> nei più svariati ambiti.

In particolare, i sondaggi che hanno valenza politica ed elettorale e che sono diffusi al pubblico, devono<sup>2</sup> essere pubblicati sul sito

[www.sondaggipoliticoelettorali.it](http://www.sondaggipoliticoelettorali.it) a cura del Dipartimento per l'Informatica e l'Editoria.

Vediamone uno: *Gli italiani ed il Referendum del 17 Aprile*

SONDAGGIO		
Dati Sondaggio	Domande	Conclusioni
Titolo del sondaggio		Soggetto che ha realizzato il sondaggio
Gli italiani ed il Referendum del 17 aprile		DEMOPOLIS - Istituto di Ricerche
Soggetto committente		Soggetto acquirente
Otto e Mezzo (LA7)		LA7 Srl
Data o periodo in cui è stato realizzato il sondaggio - Da		Data o periodo in cui è stato realizzato il sondaggio - A
22/03/2016		24/03/2016
Mezzo(i) di comunicazione di massa sul quale(i) è stato pubblicato o diffuso il sondaggio		Data di pubblicazione o diffusione
Programma Otto e Mezzo, diffuso anche da l'Espresso online e dai quotidiani Il Tirreno, Messaggero Veneto, Il Centro, La Nuova Sardegna, ecc.		30/03/2016
Popolazione di riferimento		Estensione territoriale del sondaggio
Popolazione italiana maggiorenne		Nazionale
Metodo di campionamento, inclusa l'indicazione se trattasi di campionamento probabilistico o non probabilistico, del panel e l'eventuale ponderazione		Consistenza numerica del campione di intervistati, numero dei non rispondenti e delle sostituzioni effettuate
Campione: probabilistico statisticamente rappresentativo dell'universo di riferimento, stratificato per genere, età, ampiezza demografica ed area geografica di residenza		Consistenza numerica del campione: 1.000 intervistati. Rispondenti: interviste complete effettuate 1.000 (33,1%); rifiuti/sostituzioni 2.024 (66,9%); totale contatti 3.024 (100%)
Rappresentatività del campione, inclusa l'indicazione del margine d'errore		Metodo raccolta delle informazioni
Rappresentatività dei risultati: popolazione italiana maggiorenne; margine massimo di errore 2,8%		Metodo di raccolta delle informazioni: cawi-cati per la somministrazione del questionario strutturato di rilevazione

SONDAGGIO - Domande e Risposte		
Dati Sondaggio	Domande	Conclusioni
N° Domanda	Domanda	
1	Lei sarebbe propenso a vietare alla scadenza il rinnovo delle attuali concessioni per l'estrazione in mare entro 12 miglia dalle coste?	
2	Per quali ragioni lei è favorevole al divieto di rinnovo delle concessioni?	
3	Per quali ragioni lei è contrario al divieto di rinnovo delle concessioni?	
4	Lei sa che il 17 aprile si terrà il Referendum sulle trivellazioni?	
Pagina 1 di 1 (4 elemento)		
Area Scheda Domanda e Risposta		
Testo Domanda		
Lei sarebbe propenso a vietare alla scadenza il rinnovo delle attuali concessioni per l'estrazione in mare entro 12 miglia dalle coste?		
Testo Risposta		
Sì: 74%		
No: 26%		
Dati ripercentralizzati in assenza del non sa (15%)		

Esaminiamo i termini per noi più significativi:

<sup>1</sup> I sondaggi vengono spesso fatti online mediante software quali cawi-cati, come nell'esempio riportato in figura (è indicato nella sezione "Metodo raccolta delle informazioni").

<sup>2</sup> Ciò è previsto dalla legge n. 28 del 22 Febbraio 2000, sulla par condicio.

- *Campione.* È importante avere informazioni sull'orientamento di voto di una popolazione o sull'opinione in merito ad una data questione. Un modo per ottenerlo è mediante le elezioni o i referendum; ma non si può coinvolgere l'intera popolazione tutte le volte. Pertanto, si considera un opportuno sottoinsieme della popolazione, detto appunto campione, e si registrano le preferenze espresse da esso. Le frequenze relative ottenute dal campione sono una stima delle preferenze dei singoli individui dell'intera popolazione.
- *Margine d'errore.* Se il sondaggio ha un margine massimo d'errore del 2,8%, significa che la probabilità dell'evento in questione (per esempio "l'individuo vota SI") si discosterà al massimo della quantità 0,028 dalla stima ottenuta sul campione...con probabilità "grande"<sup>3</sup>.
- *Consistenza numerica del campione.* Il numero di cittadini che costituiscono il campione è strettamente legato al margine di errore del sondaggio. Una questione significativa al riguardo è determinare la dimensione del campione affinché la stima fatta su esso sia "buona". Questa però è una questione articolata di cui non ci occupiamo.

## La questione

Una popolazione costituita da **10.000** individui è chiamata a votare tra due candidati, diciamo *A* e *B*. Mediante un sondaggio effettuato su un campione della popolazione, si è stimato che la probabilità che l'**individuo** sia favorevole ad *A* è del **40%**.

Sulla base di ciò, si vuole stimare il numero **F** di individui che voteranno *A*. Precisamente, quanto vale

$$P(3900 \leq F \leq 4100) ?$$

*Osservazioni:*

- In sostanza, a partire dalla percentuale di individui del campione che sono a favore di *A*, si vuole stimare la probabilità che il generico individuo della popolazione sia a favore di *A*.
- Gli estremi di variabilità stabiliti per *F* nel problema rappresentano rispettivamente il 39% e il 41% della popolazione. Pertanto la richiesta si può così riformulare: qual è la probabilità che il numero *F* sia compreso tra  $(40 - 1)\%$  e  $(40 + 1)\%$  della popolazione?

<sup>3</sup>Precisamente con probabilità che di solito è del 95% (in fisica, invece, si usano margini molto più bassi!). Secondo un approccio frequentista, su molti campionamenti ci si aspetta che circa il 95% delle volte, la probabilità *p* sulla popolazione cada nell'intervallo  $[\bar{p} - 0,028; \bar{p} + 0,028]$ , dove  $\bar{p}$  è la stima sul campione.

## 1.2 Un modello binomiale

Proviamo a modellizzare la situazione proposta mediante la *distribuzione binomiale*. Precisamente interpretiamo la questione nel modo seguente:

- una sequenza<sup>4</sup> di  $n = 10.000$  prove, ciascuna delle quali corrisponde ad una persona;
- ciascuna prova ha due soli esiti possibili: "la persona vota A" oppure "la persona vota B"; dall'esito del sondaggio stimiamo che la probabilità che l'individuo sia favorevole ad A è  $p = 0.4$ . Essa è costante (cioè è la stessa) per ogni prova (individuo).

Affinché abbia senso modellizzare mediante una distribuzione binomiale la v.a.  $F$ , assumiamo inoltre che le prove siano indipendenti<sup>5</sup>, ossia che il voto della persona non sia influenzato dal voto delle altre.

Secondo il nostro modello la probabilità che il numero di individui  $F$  sia, ad esempio, 3900 è<sup>6</sup>

$$P(F = 3900) = \binom{10000}{3900} \cdot 0,4^{3900} \cdot 0,6^{6100}$$

e la probabilità richiesta si può esprimere nella forma

$$P(3900 \leq F \leq 4100) = P(F = 3900) + P(F = 3901) + \dots + P(F = 4100)$$

*Osservazione.* Senza ricorrere a formule ritenute a memoria, il valore di probabilità  $P(F = 3900)$  si determina volta per volta in due passi, determinando:

1. la probabilità di una sequenza di 3900 voti a favore di  $A$  e 6100 voti a favore di  $B$ , quale

$$\underbrace{AAA\dots A}_{3900 \text{ volte}} \underbrace{BBB\dots B}_{6100 \text{ volte}}$$

per la legge della moltiplicazione (eventi indipendenti) la probabilità di tale sequenza è

$$0,4^{3900} \cdot (1 - 0,4)^{6100}$$

2. il numero di tutte le sequenze di 3900  $A$  e 6100  $B$   
esso è il numero di sottoinsiemi di 3900 elementi (posizioni per  $A$ ), contenuti in un insieme di 10000 elementi (posizioni possibili nella sequenza); dunque è

$$\binom{10000}{3900}$$

<sup>4</sup>Tale sequenza si indica spesso come "schema di Bernoulli" o "schema successo-insuccesso".

<sup>5</sup>Non possiamo dire che ciò avvenga realmente. Si tratta solamente di una nostra ipotesi che assumiamo, almeno in prima approssimazione, allo scopo di costruire un modello ragionevolmente semplice.

<sup>6</sup>Ricordiamo che la distribuzione di una variabile aleatoria binomiale  $S$  è data da

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

### C'è un problema

Il calcolo richiesto è assai articolato per la presenza dei coefficienti binomiali. Basti pensare che, ad esempio,  $70! = 1,1979 \cdot 10^{100}$ .

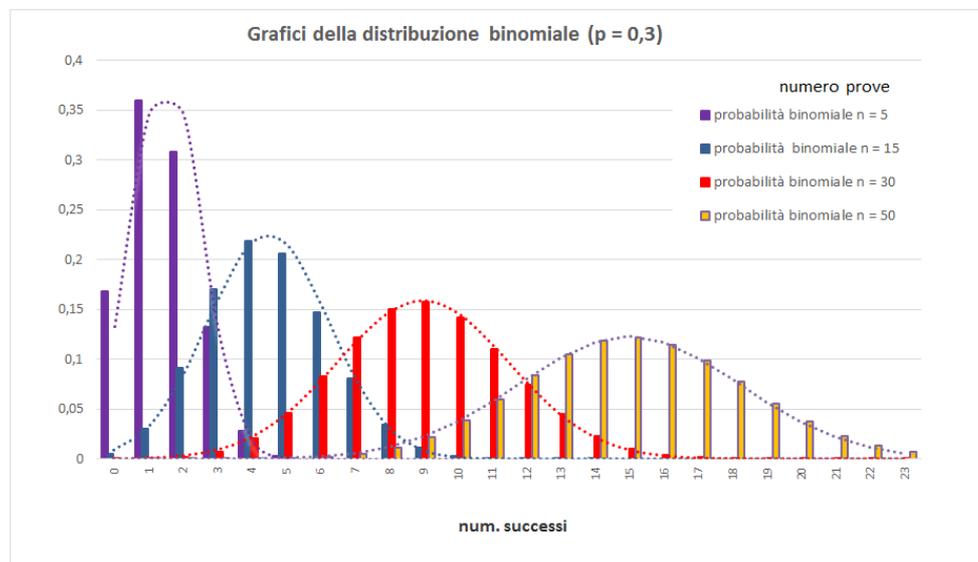
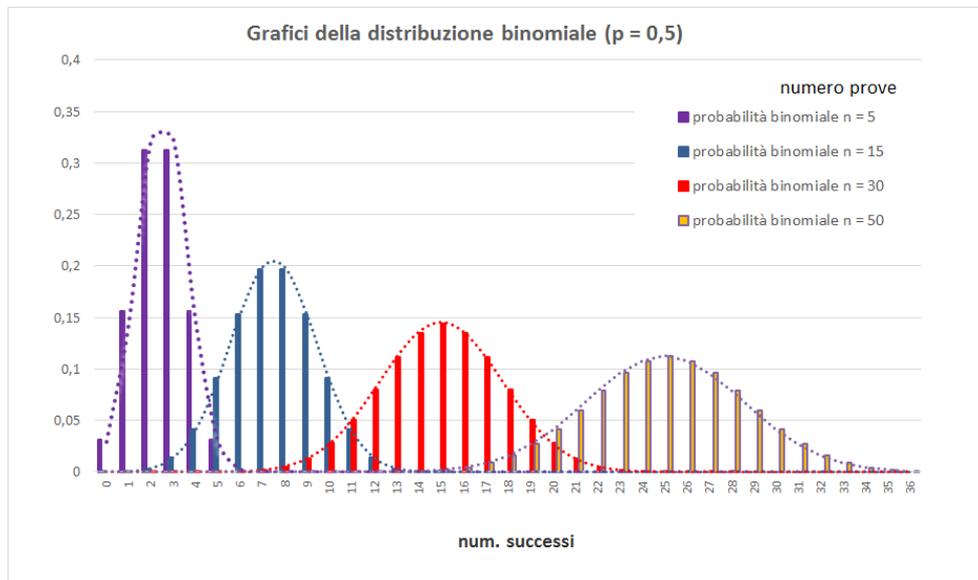
Come disse De Moivre nel Settecento, effettuare calcoli analoghi *"non è possibile senza un lavoro immenso, per non dire che è impossibile"*. E' vero che erano altri tempi e all'epoca non si disponeva dei calcolatori di oggi, ma anche noi siamo interessati a trovare un modo più efficiente per risolvere il problema.

C'è un approccio meno dispendioso computazionalmente?

### 1.3 Un nuovo modello: verso il TLC

#### L'idea

Per comprendere come risolvere il problema, seguiamo un *approccio grafico*, cioè esaminiamo i grafici della distribuzione binomiale per  $p$  fissato (dove  $p$  è la probabilità di successo della prova) al crescere del numero  $n$  di prove.



- le figure suggeriscono che il grafico della distribuzione binomiale si può *approssimare* mediante una curva che ha forma di "campana"
- tale curva è<sup>7</sup> il grafico della *densità* di una certa variabile aleatoria

<sup>7</sup>Ciò si può dimostrare formalmente, come vedremo nel paragrafo 3.1.

- e i calcoli con la "nuova" distribuzione sono *più semplici*

Pertanto:

cercheremo di approssimare la binomiale mediante la distribuzione che resta definita dalla curva "a campana".

### Un pò più precisamente

In effetti tale approssimazione è garantita da un teorema che precisa anche il senso in cui essa debba essere intesa. La sostanza di questo teorema si può formulare in prima approssimazione nel modo seguente e verrà precisata formalmente nella sezione 7.

#### TEOREMA LIMITE CENTRALE - TLC (una prima formulazione)

Sia  $S_n$  la v.a. binomiale relativa ad un numero  $n$  di prove e  $X$  una opportuna v.a. la cui densità ha come grafico la curva "a campana". Per  $n$  "grande", vale l'approssimazione:

$$\mathbf{P(a \leq S_n \leq b)} \simeq \mathbf{P(a \leq X \leq b)}$$

### Una nuova formulazione della questione

Possiamo allora usare il teorema appena visto per riformulare la nostra questione:

$$\textit{calcolare } P(3900 \leq F \leq 4100)$$

dove  $F$  è il numero di individui della popolazione a favore del candidato  $A$ .

Siamo partiti modellizzando  $F$  mediante una variabile aleatoria binomiale relativa a  $n = 10.000$  prove.

Consideriamo ora un grafico "a campana" che approssimi il grafico di tale distribuzione binomiale e indichiamo con  $\mathbf{X}$  la v.a. la cui densità ha quella "campana" come grafico. Per il TLC la nostra questione diviene

$$\textit{calcolare } \mathbf{P(3900 \leq X \leq 4100)}$$

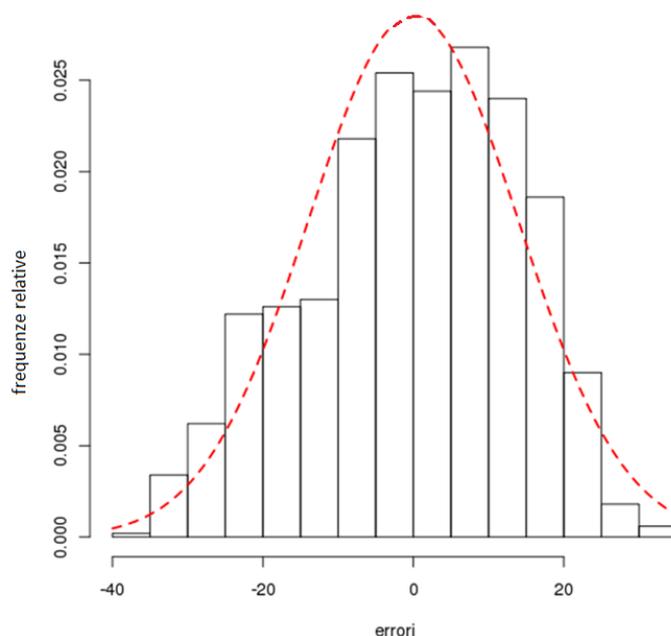
**Nota.** Questo è il nostro obiettivo; esso costituirà il filo conduttore di tutto il percorso. Vedremo, però, che sono molte le situazioni che si possono modellizzare mediante questo stesso schema. Ciò costituirà un motivo ulteriore per studiarlo a fondo.

## 1.4 Altre motivazioni alla normale

La curva "a campana" compare in molti altri contesti. Ne esaminiamo alcuni:

### Gli errori accidentali nella misura

Quando si effettua la misura di una grandezza fisica mediante uno strumento, si compiono vari errori accidentali la cui distribuzione delle frequenze relative è del tipo in figura. In opportune ipotesi, la loro distribuzione può<sup>8</sup> essere approssimata con una curva "a campana".



Significativo e coinvolgente è il seguente scritto di Galton<sup>9</sup> a tal proposito: *"Difficilmente saprei indicare cosa alcuna altrettanto adatta a colpire l'immaginazione quanto la meravigliosa **forma dell'ordine cosmico** espressa dalla Legge di frequenza degli errori. Questa legge sarebbe stata personificata dai Greci, e deificata, se ne avessero avuto conoscenza. Essa regna con serenità ed in completa indifferenza tra la confusione più selvaggia. Più è immensa la folla più è grande l'apparente anarchia, e più è perfetto il suo governo. È la **suprema legge dell'Assenza di Ragione.***

*Ogni qualvolta un grande campione di elementi caotici viene preso in mano e disposto in ordine di grandezza, una insospettata e più bella **forma di regolarità** mostra di essere stata ivi latente. Le teste della riga ordinata formano una curva che scorre con proporzioni invariabili, ed ogni elemento quando vien messo a posto, trova, come se così fosse, una nicchia predisposta, accuratamente preparata a contenerlo."*

<sup>8</sup>Da un punto di vista storico, Gauss introdusse la "nuova" distribuzione proprio per descrivere la legge degli errori casuali (1809).

<sup>9</sup>Galton F., "Order in Apparent Chaos", Natural Inheritance, n.39, 1889, p. 66; tratto da "President's Address" in Journal of Antropological Institute of Great Britain and Ireland, 15, 1886, pp. 489-499.

## Le altezze di una popolazione

Esaminiamo la statura degli italiani. Tra il 1860 e il 1905, il generale Federico Torre raccolse, per ciascuna provincia italiana, i dati relativi ad oltre ventun milioni di giovani italiani chiamati alle armi e da questi creò tavole delle frequenze relative per altezze tra i 125 e 199 cm, suddivise in intervalli di 1 cm. Come si nota in figura 1, anche questa distribuzione è molto regolare e può essere approssimata con una curva "a campana".

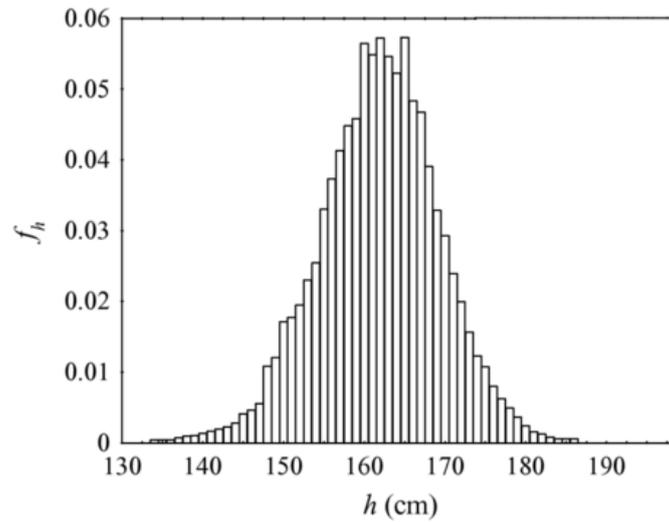


Figura 1: Distribuzione delle altezze dei circoscritti per la classe di leva 1900. La figura e i dati sono presi da [35].

## 1.5 Facciamo il punto

Abbiamo così visto come la curva "a campana" intervenga in tanti contesti. Pertanto:

- **studieremo** tale curva;
- ne esamineremo il **significato probabilistico**.

E per farlo, ci serve la sua espressione analitica.

La curva "a campana" è il grafico di una funzione della forma

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

Tali funzioni costituiscono una famiglia al variare dei parametri reali  $\mu$ ,  $\sigma$  dove  $\sigma > 0$ .

## 2 Una famiglia di funzioni: aspetti analitici

### 2.1 Prime esplorazioni -attività-

Considera la funzione  $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ . A partire dal suo grafico qualitativo, traccia su un foglio quello di

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{dove } \sigma > 0$$

per i seguenti valori dei parametri:

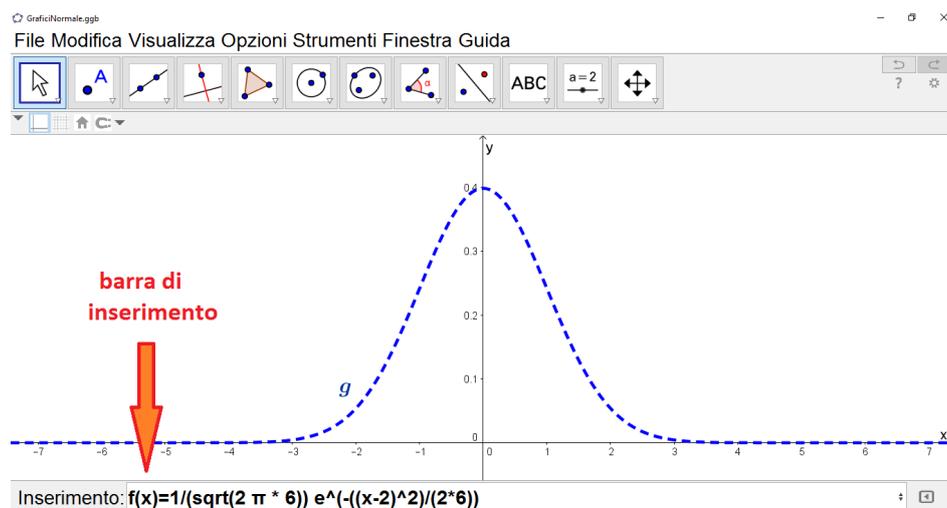
a)  $\sigma^2 = 1$  e  $\mu = +1$ ;   b)  $\sigma^2 = 4$  e  $\mu = 0$ ;   c)  $\sigma^2 = 4$  e  $\mu = -1$ .

Controlla poi i tuoi grafici mediante il file Geogebra [GraficiNormale.ggb](#)

### Suggerimento

Indichiamo con  $g$  la funzione definita da  $g(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ . Osserva che il grafico delle funzioni richieste si può ottenere mediante opportune trasformazioni del grafico di  $g$ .

### Come si utilizza il file GraficiNormale.ggb



Inserisci l'espressione analitica della funzione nella barra di inserimento. Dando l'enter viene visualizzato il grafico di tale funzione nella finestra Vista Grafici, dove puoi confrontarlo con quello già presente della funzione  $g$ .

### Risoluzione

Partendo dal grafico della funzione  $g$ , vediamo come dedurre quello della funzione  $f$ .

- a) Per  $\sigma^2 = 1$  e  $\mu = 1$  (figura 2), il grafico di  $f$  si ottiene da quello di  $g$  effettuando una traslazione lungo l'asse  $x$  di intensità 1 nel verso positivo (ossia una traslazione di vettore  $(+1, 0)$ ). Infatti la funzione  $f$  si può esprimere nella forma

$$f(x) = g(x - 1)$$

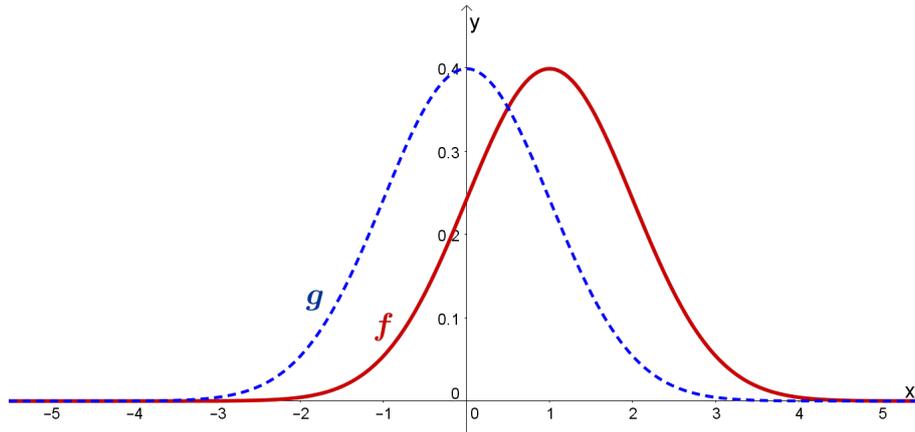


Figura 2: Grafico della densità normale  $f$  per  $\sigma^2 = 1$  e  $\mu = 1$ .

- b) Per  $\sigma^2 = 4$  e  $\mu = 0$  (figura 3), il grafico di  $f$  si ottiene dilatando quello di  $g$  del fattore  $\frac{1}{2}$  lungo l'asse  $y$  e del fattore 2 lungo l'asse  $x$ . Infatti si ha

$$f(x) = \frac{1}{2}g\left(\frac{x}{2}\right)$$

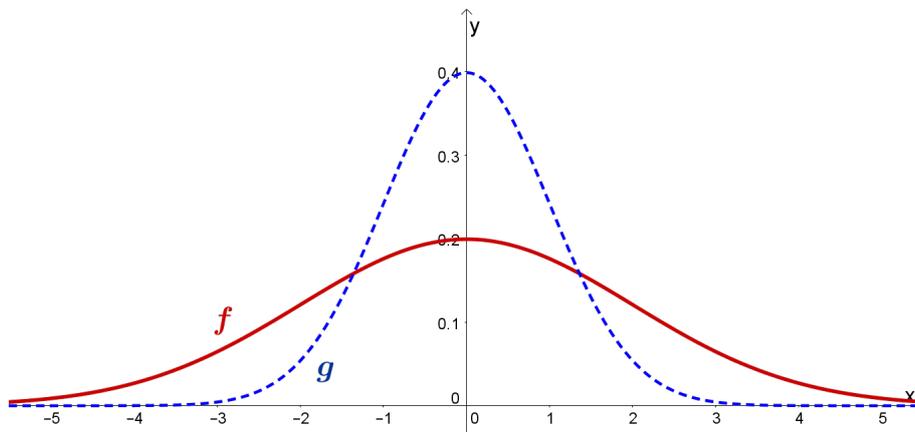


Figura 3: Grafico della densità normale  $f$  per  $\sigma^2 = 4$  e  $\mu = 0$ .

c) Nel caso  $\sigma^2 = 4$  e  $\mu = -1$  (figura 4), c'è la composizione di due trasformazioni: una del tipo esaminato nel punto b) e l'altra del tipo esaminato in a). Precisamente:

1. prima si considera la funzione  $h(x) = \frac{1}{2}g(\frac{x}{2})$  e dunque una dilatazione del grafico di  $g$  del fattore  $\frac{1}{2}$  lungo l'asse  $y$  e del fattore 2 lungo l'asse  $x$ ;
2. la funzione  $f$  si può allora esprimere come  $f(x) = h(x + 1)$  e dunque una traslazione del grafico di  $h$  lungo l'asse  $x$  di intensità 1 nel verso negativo (ossia una traslazione di vettore  $(-1, 0)$ ).

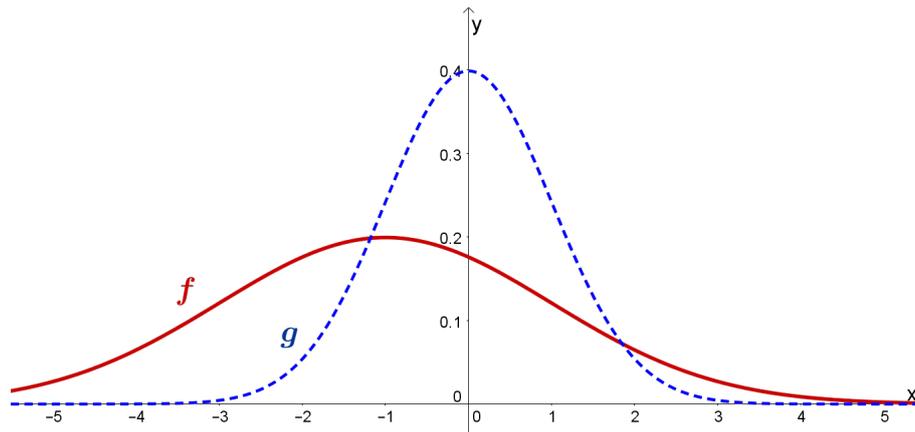


Figura 4: Grafico della densità normale  $f$  per  $\sigma^2 = 4$  e  $\mu = -1$ .

## 2.2 Caratteristiche analitiche delle funzioni -attività-

Considera la famiglia di funzioni

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

al variare dei parametri reali  $\mu$ ,  $\sigma$  dove  $\sigma > 0$ .

- a) Il grafico di  $f$  ha simmetrie?
- b) Qual è il valore massimo di  $f$  e in quale punto si ottiene?
- c) Determina i punti di flesso di  $f$ .

### Suggerimento

Eventualmente prova prima a rispondere ai quesiti relativamente a specifiche funzioni della famiglia  $f$  (esse si ottengono attribuendo ai parametri  $\mu$ ,  $\sigma$  valori specifici). Congetture le risposte dall'osservazione dei grafici, ma poi fornisci una giustificazione formale. Relativamente alle richieste a), b) non serve utilizzare la derivata.

## Risoluzione

a) Il grafico di  $f$  è simmetrico rispetto alla retta di equazione  $x = \mu$ . Ne proponiamo due dimostrazioni:

1. La funzione  $h(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}$  è pari<sup>10</sup> ossia il suo grafico è simmetrico rispetto l'asse  $y$ . Il grafico di  $f$  si ottiene da quello di  $h$  mediante la traslazione lungo l'asse  $x$  della quantità  $\mu$  e di conseguenza anche l'asse di simmetria del grafico trasla, diventando la retta  $x = \mu$ .
2. Osserviamo innanzitutto (vedi figura 5) che la condizione di simmetria rispetto alla retta  $x = \mu$ , si traduce nell'uguaglianza  $f(\mu - a) = f(\mu + a) \forall a \in \mathbb{R}$ . E in effetti vale

$$f(\mu - a) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(\mu-a-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{a^2}{2\sigma^2}} = f(\mu + a)$$

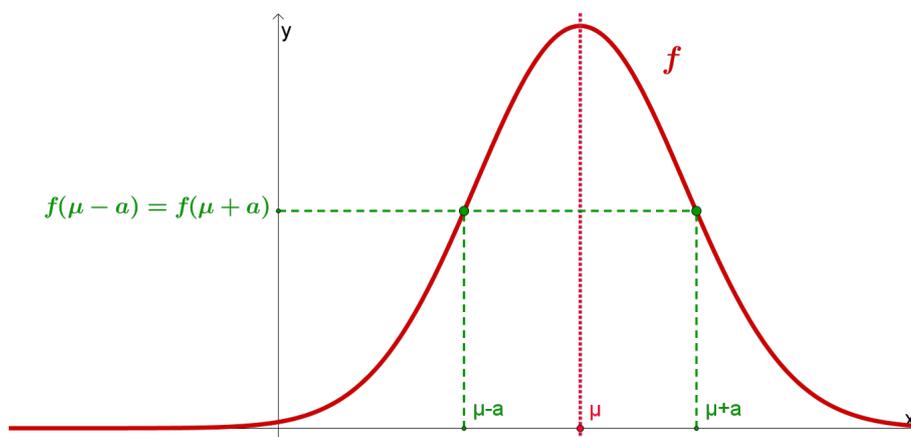


Figura 5: Simmetria della funzione  $f$ , dove si è scelto  $a > 0$ .

b) Cerchiamo innanzitutto per quale o quali valori di  $x$  la funzione  $f$  assume valore massimo. Partiamo anche in questo caso dalla funzione  $h(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}$  e osserviamo che essa assume massimo in  $x = 0$ . Per dimostrarlo basta notare che la funzione  $e^{-t}$  è decrescente e dunque lo è anche  $h$  per  $x \geq 0$ . Da qui segue che il massimo di  $h$  su  $[0; +\infty)$  si ha in  $x = 0$  e  $h$  è pari, quindi questo è anche il massimo su  $\mathbb{R}$ . Ora il grafico di  $f$  si ottiene da quello di  $h$  effettuando una traslazione lungo l'asse  $x$  della quantità  $\mu$ , per cui anche il punto di massimo trasla proprio in<sup>11</sup>  $x = \mu$ . La funzione  $f$  ha quindi valore massimo uguale a

$$f(\mu) = \frac{1}{\sqrt{2\pi\sigma}}$$

c) Per determinare i punti di flesso di  $f$ , andiamo a studiare il segno della derivata seconda. Calcoliamo quindi la derivata prima:

$$f'(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left( -\frac{2(x-\mu)}{2\sigma^2} \right) = -\frac{1}{\sqrt{2\pi\sigma^3}} (x-\mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

<sup>10</sup>Ricordiamo che una funzione  $h(x)$  definita sull'insieme  $\mathbb{R}$  è pari se  $h(x) = h(-x) \forall x \in \mathbb{R}$ .

<sup>11</sup>A questo risultato si poteva comunque arrivare anche analizzando la crescita/decrescenza della

e poi la derivata seconda:

$$\begin{aligned} f''(x) &= -\frac{1}{\sqrt{2\pi}\sigma^3} \left[ e^{-\frac{(x-\mu)^2}{2\sigma^2}} + (x-\mu)e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left( -\frac{x-\mu}{\sigma^2} \right) \right] \\ &= \frac{1}{\sqrt{2\pi}\sigma^5} e^{-\frac{(x-\mu)^2}{2\sigma^2}} ((x-\mu)^2 - \sigma^2) \end{aligned}$$

Il segno di  $f''(x)$  dipende solo dal fattore  $(x-\mu)^2 - \sigma^2$ , dato che la funzione esponenziale assume sempre valori positivi e  $\sigma$  è positivo per ipotesi. Per studiare il segno della funzione  $(x-\mu)^2 - \sigma^2$ , troviamo prima i suoi zeri<sup>12</sup>, che sono  $x = \mu - \sigma$  e  $x = \mu + \sigma$ . Essa è una funzione polinomiale di secondo grado e il coefficiente del termine  $x^2$  è positivo, perciò possiamo schematizzare il suo segno con lo schema seguente:

$$\begin{array}{cccc} f'' & + + + & - - - & + + + \\ f & \cap & \cup & \cap \end{array}$$

La funzione  $f$  ha quindi flessi nei punti

$$x = \mu - \sigma \vee x = \mu + \sigma$$

Osserviamo che tali punti sono simmetrici rispetto al punto  $x = \mu$ , in accordo con il fatto che il grafico di  $f$  è simmetrico rispetto alla retta  $x = \mu$  (un grafico riassuntivo è la figura del paragrafo 2.3).

---

funzione tramite lo studio del segno della derivata prima.

<sup>12</sup> $(x-\mu)^2 - \sigma^2 = 0 \iff (x-\mu)^2 = \sigma^2 \iff x-\mu = +\sigma \vee x-\mu = -\sigma.$

## 2.3 Significato geometrico dei parametri -attività-

Considera la famiglia di funzioni

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

al variare dei parametri reali  $\mu$ ,  $\sigma$  dove  $\sigma > 0$ .

a) Qual è il significato geometrico di  $\mu$ ?

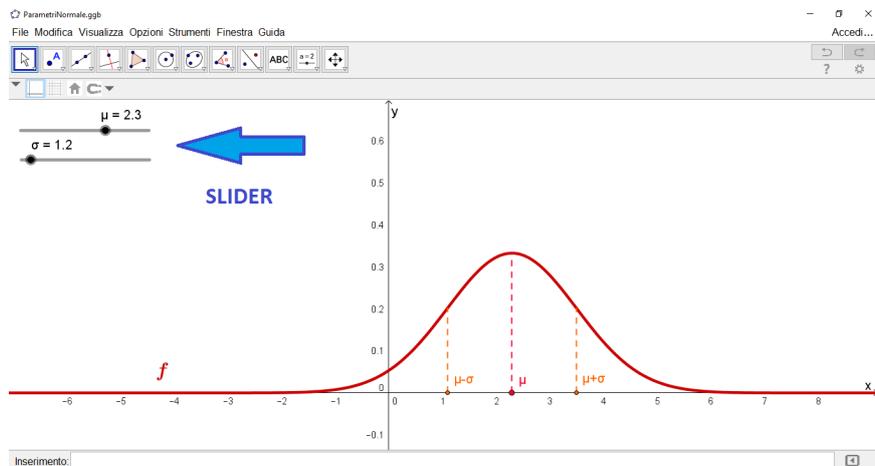
b) Qual è il significato geometrico di  $\sigma^2$ ?

Ossia spiega come varia il grafico della funzione  $f$  al variare dei due parametri  $\mu$  e  $\sigma$ .

### Suggerimento

Fissa un valore per un parametro e descrivi come variano i grafici al variare dell'altro parametro. Effettua delle congetture mediante il file Geogebra [ParametriNormale.ggb](#) e poi giustificalo formalmente.

### Come si utilizza il file ParametriNormale.ggb



Il file dispone di due slider, ciascuno relativo ad un parametro della famiglia. Muovendo ciascun cursore variano i valori che si attribuiscono al corrispondente parametro; e dunque variano anche i grafici delle funzioni individuate da tali valori del parametro.

### Risoluzione

a) Fissiamo per  $\sigma^2$  il valore<sup>13</sup> 1. Al variare di  $\mu$  il grafico di  $f$  *trasla* lungo l'asse  $x$  (come suggerisce la figura 6). Partiamo, infatti, dal grafico di

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

<sup>13</sup>Per qualsiasi altro valore di  $\sigma^2$  si giunge alle stesse conclusioni.

Allora

$$f(x) = g(x - \mu)$$

quindi il grafico di  $f$  si ottiene da quello di  $g$  mediante una traslazione lungo l'asse  $x$  di intensità  $|\mu|$  nel verso positivo se  $\mu > 0$ , nel verso negativo se  $\mu < 0$ .

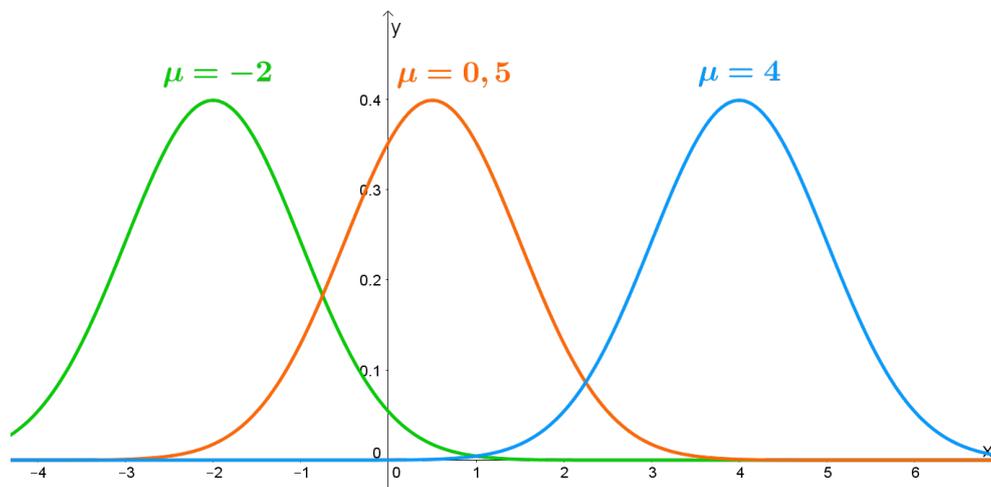


Figura 6: Grafici di  $f$  per  $\sigma^2 = 1$  e più valori di  $\mu$ .

- b) Fissiamo per  $\mu$  il valore<sup>14</sup> 0. Al variare di  $\sigma^2$  il grafico di  $f$  si *dilata* (come suggerito in figura 7). Più precisamente, al crescere di  $\sigma^2$ , *cresce* l'apertura del grafico e *diminuisce* il massimo della funzione. Se partiamo, infatti, dal grafico di

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

allora

$$f(x) = \frac{1}{\sigma} g\left(\frac{x}{\sigma}\right)$$

Sull'asse delle  $x$  abbiamo una dilatazione di rapporto  $\sigma$ , perciò al crescere di  $\sigma^2$  la curva si allarga orizzontalmente. Sulle ordinate abbiamo invece una dilatazione del fattore  $\frac{1}{\sigma}$  per cui al crescere di  $\sigma^2$  il grafico di  $f$  si contrae verticalmente.

---

<sup>14</sup>Per qualsiasi altro valore di  $\mu$  si giunge alle stesse conclusioni.

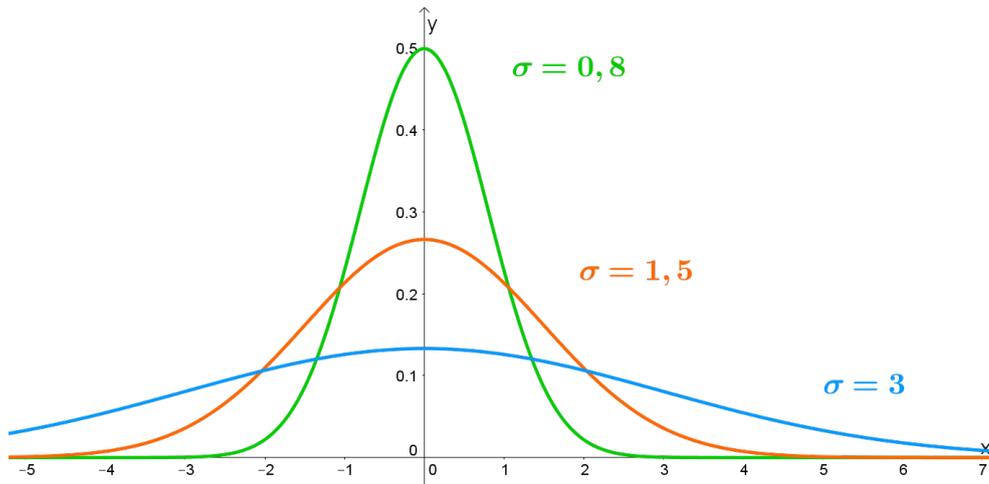


Figura 7: Grafici di  $f$  per  $\mu = 0$  e più valori di  $\sigma^2$ .

## 2.4 Conclusioni

Consideriamo la famiglia di funzioni

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{dove } \sigma > 0$$

al variare dei parametri reali  $\mu$  e  $\sigma$ .

### Caratteristiche delle funzioni:

- il grafico di  $f$  è *simmetrico* rispetto alla retta di equazione  $x = \mu$
- la funzione  $f$  assume valore *massimo* per  $x = \mu$ ;  
esso è  $f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$
- la funzione  $f$  ha *flessi* nei punti  $x = \mu - \sigma$  e  $x = \mu + \sigma$

### Interpretazione geometrica dei parametri:

- al variare di  $\mu$  il grafico di  $f$  *trasla* lungo l'asse  $x$
- al variare di  $\sigma^2$  il grafico di  $f$  si *dilata*. Più precisamente, al crescere di  $\sigma^2$ , cresce l'apertura del grafico e diminuisce il massimo della funzione

### 3 Una famiglia di funzioni: interpretazione probabilistica

#### 3.1 Funzione di densità

La funzione

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

è una funzione di densità per ogni  $\mu, \sigma \in \mathbb{R}$ , con  $\sigma > 0$ .

Infatti vale:

- $f(x) \geq 0 \forall x \in \mathbb{R}$   
dato che la funzione esponenziale assume solo valori positivi e  $\sigma$  è positivo per ipotesi<sup>a</sup>

- $\int_{-\infty}^{+\infty} f(x) dx = 1$

Si può dimostrare<sup>b</sup> che  $\int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sqrt{2\pi}\sigma$ . Pertanto il fattore  $\frac{1}{\sqrt{2\pi}\sigma}$  che compare nella funzione  $f$  serve proprio affinché l'integrale di  $f$  su  $\mathbb{R}$  sia 1

**Definizione.** La funzione  $f$  si dice funzione di **densità normale** (con parametri  $\mu$  e  $\sigma^2$ ) e la variabile aleatoria (spesso abbreviata in v.a.) ad essa relativa si dice **variabile aleatoria normale**.

---

<sup>a</sup>Anzi  $f(x) > 0 \forall x \in \mathbb{R}$ .

<sup>b</sup>Con strumenti di analisi che vanno oltre le competenze liceali.

*Osservazione.* La probabilità che  $X$  sia compresa tra due valori  $a, b \in \mathbb{R}$ , come per ogni v.a. continua, si può esprimere nella forma:

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

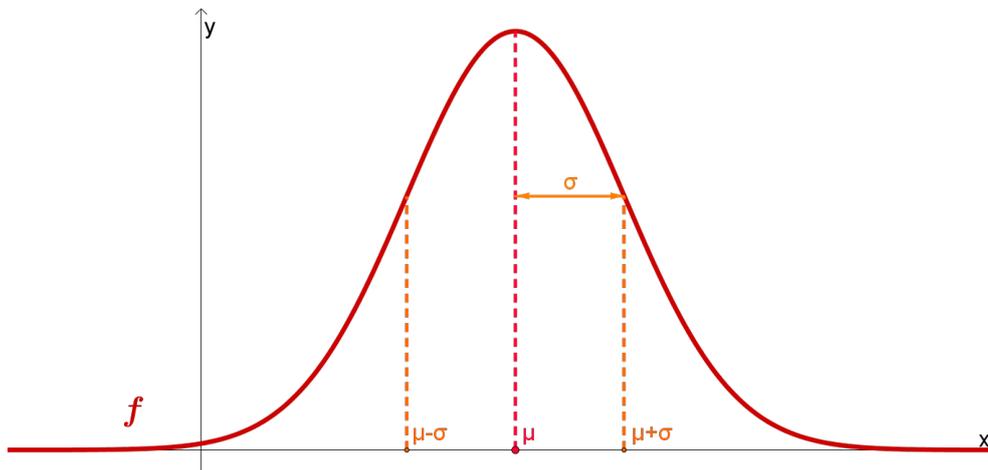
dove  $f$  è la funzione di densità normale.

C'è però un problema: tale integrale *non* è risolvibile elementarmente, cioè la primitiva di  $f$  non si può esprimere mediante una formula che contenga le funzioni elementari e le operazioni usuali. Vedremo nel seguito come risolvere tale questione.

### 3.2 Significato probabilistico dei parametri

Per comprendere il significato probabilistico dei parametri  $\mu$  e  $\sigma$ , iniziamo con l'esaminare l'espressione analitica e il grafico della densità:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{dove } \sigma > 0$$



Da essi deduciamo il significato geometrico dei due parametri, come visto nel paragrafo precedente (2.3).

Ora però compiamo un ulteriore passo: utilizziamo il fatto che  $f$  è una densità di *probabilità*. Queste due considerazioni<sup>15</sup> insieme, cosa permettono di affermare sui parametri?

*Intuitivamente* possiamo dire che:

- $\mu$  rappresenta il "centro" della distribuzione
- $\sigma^2$  è una misura della "larghezza" della distribuzione, rispetto al "centro"

Ci *aspettiamo* dunque che:

- $\mu$  sia uguale o legato al valore atteso della v.a. normale
- $\sigma^2$  sia uguale o legato alla varianza della v.a. normale

Precisamente, mediante il calcolo integrale si può dimostrare il seguente risultato:

**Proposizione.** Sia  $X$  una v.a. normale, cioè una v.a. continua che ha densità

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{dove } \sigma > 0$$

Allora la media di  $X$  è  $\mu$  e la varianza di  $X$  è  $\sigma^2$ .

Proveremo questo risultato nel caso della v.a. normale di parametri  $\mu = 0$  e  $\sigma^2 = 1$ . Tale variabile si indica spesso con  $Z$  e si dice v.a. **normale standard**.

<sup>15</sup>Ossia il significato geometrico dei parametri e il fatto che  $f$  è una densità di probabilità.

### 3.3 Media e varianza della distribuzione normale standard -attività-

Consideriamo la variabile aleatoria normale standard, cioè la v.a. continua che ha densità

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

Mostra che la media di tale variabile è 0 e la varianza è 1.

#### Suggerimenti

- a) Ricorda che, per definizione di media, la media della v.a. normale standard è

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

Per determinare il valore di questo integrale non serve scrivere esplicitamente una primitiva della funzione  $x f(x)$ : basta osservare che essa è dispari.

- b) Secondo la definizione di varianza, quella della v.a. normale standard risulta uguale a

$$\int_{-\infty}^{+\infty} x^2 f(x) dx$$

Per determinare il valore di questo integrale, osserva che la funzione  $x^2 e^{-x^2/2}$  si può scrivere nella forma  $x(xe^{-x^2/2})$  e dunque si può integrare per parti.

## Risoluzione

- a) Per trovare la media della v.a. normale standard calcoliamo dunque il seguente integrale:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-x^2/2} dx$$

Osserviamo che la funzione  $g(x) = x e^{-x^2/2}$  è dispari, ossia vale  $g(-x) = -g(x) \forall x \in \mathbb{R}$ . Ciò significa che il grafico della funzione  $g$  è simmetrico rispetto l'origine degli assi. Tale proprietà ci permette di dedurre direttamente il valore dell'integrale in esame.

Per comprenderlo vediamo una situazione più semplice e consideriamo una funzione  $h$  dispari, come quella rappresentata in figura 8. Per ragioni di simmetria, l'area del sottoinsieme del piano delimitato dal grafico di  $h$  e dall'asse  $x$  nell'intervallo  $[-t; 0]$  è uguale all'area del sottoinsieme del piano delimitato dal grafico di  $h$  e dall'asse  $x$  nell'intervallo  $[0; t]$ , per qualsiasi reale  $t > 0$ . Ma la funzione  $h$  assume segni opposti in questi due intervalli, per cui

$$\int_{-t}^0 h(x) dx = - \int_0^t h(x) dx$$

Pertanto vale anche

$$\int_{-\infty}^0 h(x) dx = - \int_0^{+\infty} h(x) dx$$

e quindi

$$\int_{-\infty}^{+\infty} h(x) dx = 0$$

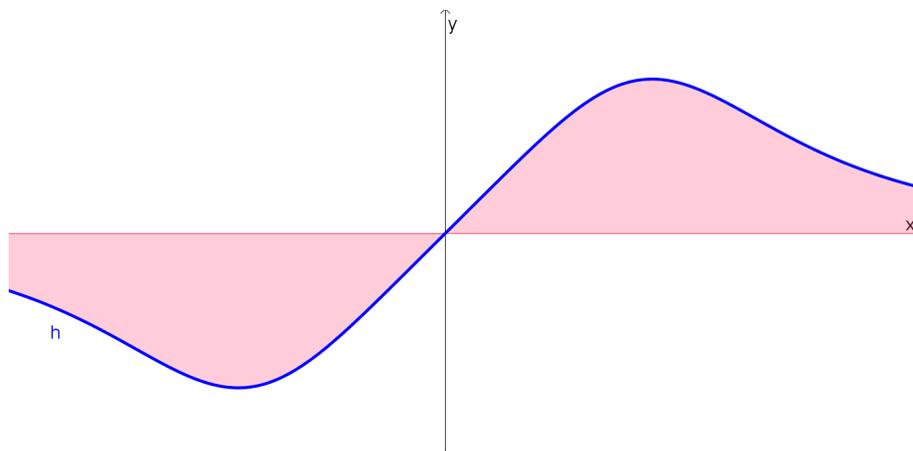


Figura 8: Rappresentazione del sottoinsieme del piano delimitato dall'asse  $x$  e dal grafico della funzione  $h(x) = \frac{x}{1+x^4}$ .

Più in generale, ragionando come nell'esempio, si ottiene che l'integrale su  $\mathbb{R}$  di una funzione (integrabile) dispari è 0. Per cui anche nel nostro caso

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-x^2/2} dx = 0$$

Ossia la media della v.a. normale standard è 0.

b) Vediamo ora il calcolo della varianza della v.a. normale standard. Nel nostro caso essa è uguale a

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-x^2/2} dx$$

Integriamo per parti:

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \left( x e^{-x^2/2} \right) dx &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \left( -e^{-x^2/2} \right)' dx \\ &= -\frac{1}{\sqrt{2\pi}} \left[ x e^{-x^2/2} \right]_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx \\ &= 0 + 1 = 1 \end{aligned}$$

Il primo addendo è 0: essenzialmente ciò è dovuto al fatto che<sup>16</sup>

$\lim_{x \rightarrow +\infty} (x e^{-x^2/2}) = 0$ . Per determinare il secondo addendo, invece, utilizziamo il fatto che  $f$  è una densità (si veda paragrafo 3.1) e dunque vale

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

Concludiamo così che la varianza della v.a. normale standard è 1.

---

<sup>16</sup>Precisamente:

$$\left[ x e^{-x^2/2} \right]_{-\infty}^{+\infty} = \lim_{k \rightarrow -\infty} \left[ x e^{-x^2/2} \right]_k^0 + \lim_{h \rightarrow +\infty} \left[ x e^{-x^2/2} \right]_0^h = \lim_{k \rightarrow -\infty} (-k e^{-k^2/2}) + \lim_{h \rightarrow +\infty} (h e^{-h^2/2}) = 0 + 0$$

dove con la prima uguaglianza abbiamo esplicitato il significato della prima scrittura.

## 4 Standardizzazione

Abbiamo visto che, per risolvere il nostro problema guida relativo ai sondaggi (paragrafo 1.3), conviene ricondursi a calcolare probabilità relative alla distribuzione normale. Come fare? È quanto vedremo in questa sezione.

### 4.1 Standardizzazione della variabile aleatoria normale -video-

*In questo video ti viene illustrato come esprimere probabilità relative ad una variabile aleatoria normale in termini di probabilità relative ad una variabile aleatoria speciale: la variabile aleatoria standard. Ti sono indicati anche i motivi sottesi ai vari passi, in modo che tu possa ricostruire il procedimento anche a distanza di tempo.*



Il video si trova all'indirizzo <https://youtu.be/M7Kt39446sU>

## 4.2 Calcolo di probabilità relative alla distribuzione normale standard -video-

*In questo video ti viene illustrato come calcolare probabilità relative alla variabile aleatoria normale standard, utilizzando opportune tavole. Questo video e il video 4.1, che ti consigliamo di esaminare, ti forniscono gli strumenti per calcolare probabilità relative alla distribuzione normale. Ti proponiamo anche una loro sintesi scritta nel paragrafo 4.3.*



Il video si trova all'indirizzo <https://youtu.be/Ywtt4tIZXVo>

### 4.3 Calcolo di probabilità relative alla distribuzione normale

Tale procedimento è illustrato in dettaglio in due video (paragrafi 4.1 e 4.2) che ti consigliamo di esaminare. Ne proponiamo qui una sintesi.

Consideriamo una generica v.a. normale  $X$  di media  $\mu$  e varianza  $\sigma^2$ . Dati due valori reali  $a, b$  con  $a < b$ , vogliamo calcolare

$$P(a \leq X \leq b)$$

Come visto nell'osservazione del paragrafo 3.1, tale probabilità non si può però calcolare mediante le formule "usuali".

Un modo per risolvere la questione è ricorrere a tavole che contengono valori (approssimati) di probabilità. Più precisamente essi sono relativi alla v.a. normale con media 0 e varianza 1, cioè alla v.a. standard  $Z$ . Per poter usare le tavole occorre dunque prima passare da valori di  $X$  a valori di  $Z$ .

Come fare? Proviamo a ragionare in termini di variabili aleatorie:

1. per avere *media*<sup>a</sup> 0, trasformiamo  $X$  nella nuova variabile  $X - \mu$
2. per avere *varianza*<sup>b</sup> 1, trasformiamo  $X - \mu$  nella nuova variabile  $\frac{X - \mu}{\sigma}$

<sup>a</sup>Come la media della v.a.  $Z$ .

<sup>b</sup>Come la varianza della v.a.  $Z$ .

Pertanto, ogni v.a. normale  $X$  si può ricondurre<sup>17</sup> alla v.a. standard  $Z$  mediante la trasformazione

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

Per risolvere il problema iniziale consideriamo dunque i valori trasformati di  $a$  e  $b$  secondo tale relazione, ossia

$$a' = \frac{a - \mu}{\sigma} \quad \text{e} \quad b' = \frac{b - \mu}{\sigma}$$

Allora la probabilità richiesta si può così esprimere<sup>a</sup>:

$$P(a \leq X \leq b) = P(a' \leq Z \leq b')$$

<sup>a</sup>Tale risultato si può intuire per la costruzione fatta e si può comunque dimostrare formalmente (paragrafo 4.5).

<sup>17</sup>Nel paragrafo 4.4 discutiamo una dimostrazione formale dei due fatti enunciati. Ossia che la media e la varianza di  $\frac{X - \mu}{\sigma}$  sono rispettivamente 0 e 1.

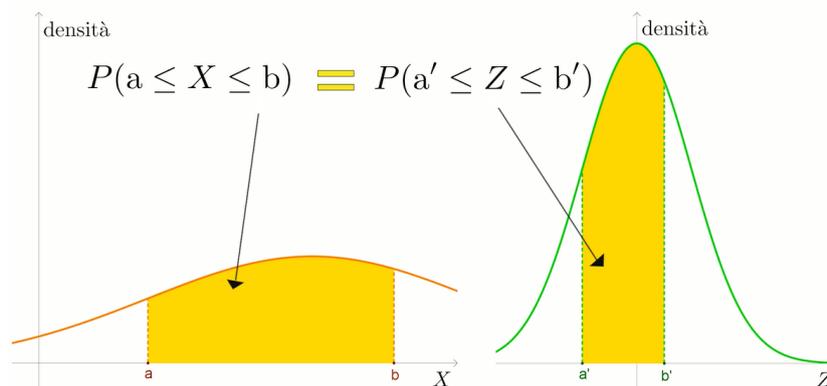


Figura 9: Vediamo qui un'interpretazione grafica, mediante le aree, dell'uguaglianza  $P(a \leq X \leq b) = P(a' \leq Z \leq b')$ .

Resta da determinare il valore di  $P(a' \leq Z \leq b')$ .

Possiamo ora utilizzare le tavole. Esse contengono valori approssimati di probabilità del tipo

$$P(Z \leq k) \quad \text{dove } k \text{ è un valore positivo}$$

Per usare le tavole, si procede come suggerisce la figura 10 per l'esempio  $P(Z \leq 1,23)$ .

Se invece occorre calcolare probabilità su altri intervalli, l'idea è di esprimere quanto richiesto in termini di probabilità del tipo<sup>18</sup>  $P(Z \leq k)$  con  $k$  positivo. Per far ciò, si possono utilizzare la simmetria della densità normale  $f$  e le proprietà degli integrali, interpretandole magari sul grafico di  $f$  in termini di *aree*.

Ad esempio, per calcolare  $P(0,98 \leq Z \leq 1,74)$ , che è una probabilità della forma in esame, si può procedere come illustrato in figura 11.

<sup>18</sup>Dato che quelle sono le probabilità che si trovano sulle tavole.

$z$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Figura 10: Lettura sulla tavola del valore approssimato di probabilità per  $k = 1, 23$ . Quindi  $P(Z \leq 1, 23) \simeq 0, 8907$ .

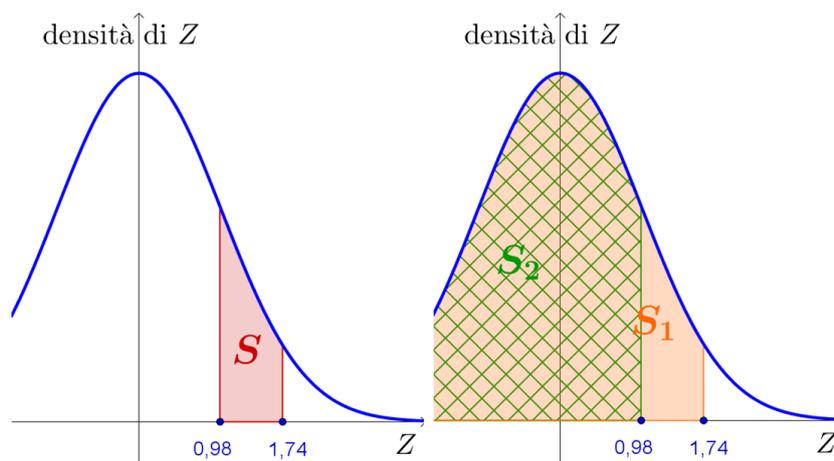


Figura 11:  $\text{Area}(S) = \text{Area}(S_1) - \text{Area}(S_2) \implies P(0,98 \leq Z \leq 1,74) = P(Z \leq 1,74) - P(Z \leq 0,98)$ .

#### 4.4 Media e varianza della variabile aleatoria standardizzata -attività-

Sia  $X$  la variabile aleatoria normale di media  $\mu$  e varianza  $\sigma^2$ .  
Dimostra che la variabile aleatoria  $\frac{X-\mu}{\sigma}$  ha media 0 e varianza 1.

**Nota.** Questo ci permette di concludere che  $\frac{X-\mu}{\sigma}$  è<sup>19</sup> una variabile aleatoria con distribuzione normale standard.

#### Suggerimento

Usa le proprietà algebriche della media (varianza) per esprimere la media (varianza) di  $\frac{X-\mu}{\sigma}$  in termini della media (varianza) della variabile  $X$ , dato che questa è nota. In particolare ricorda che per ogni numero  $c$  vale

$$\begin{aligned} E(cX) &= cE(X) & V(cX) &= c^2V(X) \\ E(X+c) &= E(X)+c & V(c+X) &= V(X) \end{aligned}$$

Con le notazioni  $E(Y)$ ,  $V(Y)$  indichiamo, come si usa spesso, rispettivamente la media e la varianza della v.a.  $Y$ .

---

<sup>19</sup>A rigore, per concludere che la variabile aleatoria così trasformata ha distribuzione normale standard, serve qualche considerazione aggiuntiva. Ma preferiamo non occuparcene.

### Risoluzione

Partiamo col calcolare  $E\left(\frac{X-\mu}{\sigma}\right)$ . Sappiamo che  $E(X) = \mu$ , perciò cerchiamo di ricondurci alla variabile  $X$ , usando le proprietà algebriche della varianza:

$$E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E(X-\mu) = \frac{1}{\sigma}(E(X)-\mu) = \frac{1}{\sigma}(\mu-\mu) = 0$$

Abbiamo così provato che la media cercata è **0**.

Calcoliamo ora la  $V\left(\frac{X-\mu}{\sigma}\right)$ . Come prima cerchiamo di ricondurci alla variabile  $X$  visto che di questa sappiamo che la varianza è  $\sigma^2$ . Usando le proprietà algebriche della varianza si ha

$$V\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}V(X-\mu) = \frac{1}{\sigma^2}V(X) = \frac{1}{\sigma^2}\sigma^2 = 1$$

Abbiamo così provato che la varianza cercata è **1**.

#### 4.5 Probabilità con la variabile aleatoria standardizzata -attività-

Sia  $X$  la variabile aleatoria normale di media  $\mu$  e varianza  $\sigma^2$ . Dimostra che vale

$$P(a \leq X \leq b) = P(a' \leq Z \leq b')$$

dove  $Z$  è la variabile aleatoria normale standard e  $a'$  e  $b'$  sono i valori standardizzati<sup>a</sup> di  $a$  e  $b$ .

---

<sup>a</sup>Ossia  $a' = \frac{a-\mu}{\sigma}$  e  $b' = \frac{b-\mu}{\sigma}$ .

#### Suggerimento

Scrivi esplicitamente le due probabilità  $P(a \leq X \leq b)$ ,  $P(a' \leq Z \leq b')$  come integrali delle densità delle rispettive variabili. Poi per provare l'uguaglianza richiesta, conviene utilizzare una opportuna sostituzione nell'integrale relativo alla v.a.  $X$ .

### Risoluzione

Iniziamo esprimendo  $P(a \leq X \leq b)$  come integrale della funzione densità di  $X$ , cioè

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Analogamente

$$P(a' \leq Z \leq b') = \int_{a'}^{b'} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

Dobbiamo provare l'uguaglianza tra questi due integrali. Osservando gli esponenti nei due integrali oppure i loro estremi di integrazione, ci accorgiamo che conviene effettuare la sostituzione

$$z = \frac{x - \mu}{\sigma}$$

D'altronde si usa la stessa trasformazione per passare dalla v.a.  $X$  alla v.a.  $Z$ . Pertanto

$$x = \sigma z + \mu \quad \implies \quad dx = \sigma dz$$

Con tale sostituzione si ha dunque

$$\begin{aligned} P(a \leq X \leq b) &= \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2}} \sigma dz \\ &= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \end{aligned}$$

Ma, come prima ricordato,

$$\int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$$

Abbiamo così provato formalmente che

$$P(a \leq X \leq b) = P(a' \leq Z \leq b')$$

dove

$$a' = \frac{a - \mu}{\sigma} \quad \text{e} \quad b' = \frac{b - \mu}{\sigma}$$

## 5 Aspetti di calcolo

### 5.1 Un esempio di calcolo -attività-

Il calcolo di probabilità relative alla v.a. normale è illustrato in dettaglio in due video (paragrafi 4.1 e 4.2) che ti consigliamo di esaminare prima di risolvere l'esercizio.

Sia  $X$  la v.a. normale di media  $\mu = 5$  e varianza  $\sigma^2 = 9$ .  
Calcola  $P(3,5 \leq X \leq 11)$ , ossia la probabilità che  $X$  sia compresa tra 3,5 e 11.

#### Risoluzione

Procediamo seguendo il ragionamento visto dettagliatamente nei video e in sintesi nel paragrafo 4.3:

1. Standardizziamo la v.a.  $X$  mediante la trasformazione  $Z = \frac{X-\mu}{\sigma}$ . Pertanto i nuovi estremi di variabilità di  $Z$  diventano<sup>20</sup>  $-0,5$  e  $2$  e vale

$$P(3,5 \leq X \leq 11) = P(-0,5 \leq Z \leq 2)$$

2. Per determinare la probabilità in  $Z$ , ricordiamo che le tavole forniscono valori della forma  $P(Z \leq k)$  con  $k$  positivo. Perciò esprimiamo  $P(-0,5 \leq Z \leq 2)$  in termini di probabilità di tale forma:

- per l'additività dell'integrale sul dominio di integrazione<sup>21</sup>

$$\mathbf{P(-0,5 \leq Z \leq 2)} = P(Z \leq 2) - P(Z \leq -0,5)$$

- per la simmetria del grafico della densità normale standard<sup>22</sup>

$$P(Z \leq -0,5) = P(Z \geq 0,5) = 1 - P(Z \leq 0,5)$$

In sintesi le ultime uguaglianze ci dicono che

$$\mathbf{P(-0,5 \leq Z \leq 2)} = P(Z \leq 2) - 1 + P(Z \leq 0,5)$$

3. Andiamo ora a leggere sulle tavole i valori di probabilità richiesti

$$P(Z \leq 2) \simeq 0,9772 \quad \text{e} \quad P(Z \leq 0,5) \simeq 0,6915$$

Possiamo così concludere che

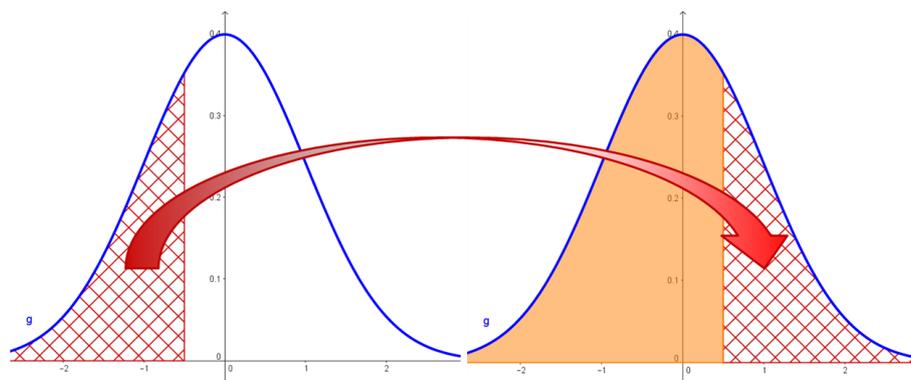
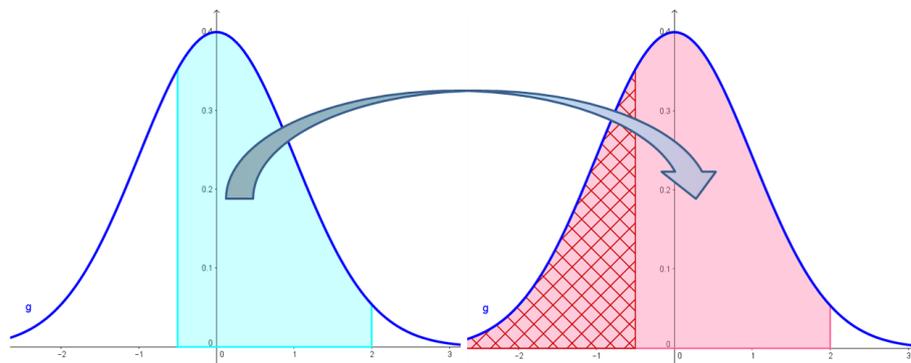
$$\begin{aligned} P(3,5 \leq X \leq 11) &\simeq 0,9772 - 1 + 0,6915 \\ &\simeq \boxed{0,67} \end{aligned}$$

<sup>20</sup>Usando le notazioni del paragrafo 4.3 e dei video, stiamo dicendo che  $a = 3,5$   $b = 11$  e  $a' = -0,5$   $b' = 2$ .

<sup>21</sup>Precisamente  $\int_{-0,5}^2 g(x) dx = \int_{-\infty}^2 g(x) dx - \int_{-\infty}^{-0,5} g(x) dx$  dove  $g$  è la densità di  $Z$ .

<sup>22</sup>Nella seconda uguaglianza passiamo all'insieme complementare.

*Osservazione.* Nel punto 2. abbiamo fatto ricorso alle proprietà degli integrali e alla simmetria del grafico della funzione densità normale  $g$ . E' espressivo (e d'aiuto) interpretare tali probabilità in termini di *aree* sul grafico di  $g$ .



## 5.2 Valori di probabilità notevoli -attività-

Di solito le tavole della normale standard non riportano valori di  $Z$  superiori<sup>23</sup> a 4. Perché? L'attività seguente suggerirà la risposta.

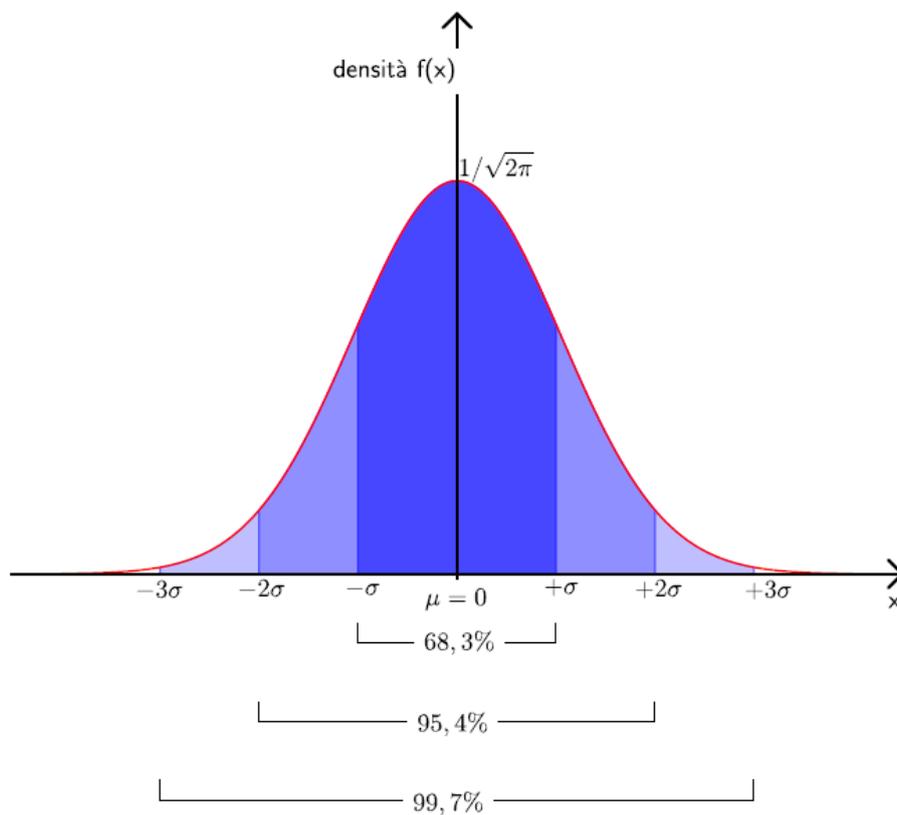
Sia  $X$  una variabile aleatoria normale con parametri  $\mu$  e  $\sigma^2$ .  
Verifica che vale

$$P(\mu - \sigma \leq X \leq \mu + \sigma) \simeq 68,3\%$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \simeq 95,4\%$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \simeq 99,7\%$$

$$P(\mu - 4\sigma \leq X \leq \mu + 4\sigma) \simeq 100\%$$



<sup>23</sup>Anzi, le tavole considerate nei video non superano il valore 3.

## Risoluzione

Mostriamo per esempio<sup>24</sup> la prima approssimazione. Vogliamo valutare la probabilità che la variabile  $X$  si discosti dalla media  $\mu$  di al più una quantità pari a  $\sigma$ .

Per determinare questo valore di probabilità seguiamo il procedimento mostrato nel paragrafo 4.3:

1. standardizziamo la variabile  $X$

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = P(-1 \leq Z \leq 1)$$

2. per la simmetria del grafico della densità normale e per le proprietà degli integrali si ha

$$P(-1 \leq Z \leq 1) = 2P(0 \leq Z \leq 1) = 2(P(Z \leq 1) - P(Z \leq 0))$$

3. ricorrendo alle tavole, otteniamo

$$2 \cdot (P(Z \leq 1) - P(Z \leq 0)) \simeq 2 \cdot (0,8413 - 0,5) \simeq \boxed{0,683}$$

In conclusione, la probabilità che la v.a. normale  $X$  assuma valori tra  $\mu - \sigma$  e  $\mu + \sigma$  è quasi il 70%.

---

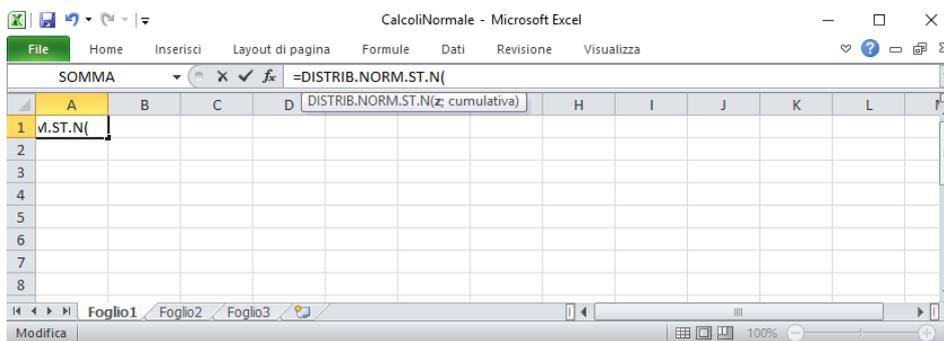
<sup>24</sup>Analogamente è possibile calcolare gli altri valori di probabilità notevoli.

### 5.3 Valori di probabilità mediante il foglio elettronico

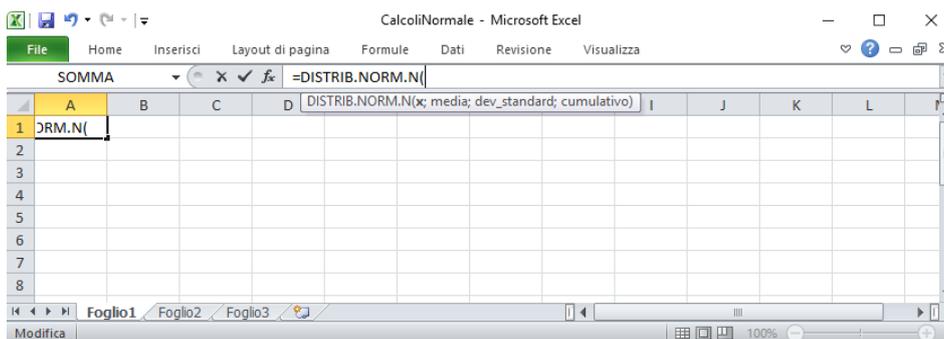
Oltre alle tavole, un altro possibile strumento a cui ricorrere per determinare valori (approssimati) di probabilità è il foglio elettronico, ad esempio Excel.

Su **Excel** sono presenti infatti:

- la funzione *DISTRIB.NORM.ST.N*( $z$ ; *cumulativa*);  
se *cumulativo*=VERO, la funzione restituisce il valore di probabilità  $P(Z \leq z)$ , dove  $Z$  è la v.a. normale standard;  
se *cumulativo*=FALSO, dà il valore  $g(z)$  dove  $g$  è la densità di  $Z$ .



- la funzione *DISTRIB.NORM.N*( $x$ ; *media*; *dev\_standard*; *cumulativo*)  
se *cumulativo*=VERO, la funzione restituisce il valore di probabilità  $P(X \leq x)$  dove  $X$  è la v.a. normale con  $\mu = \text{media}$  e  $\sigma = \text{dev\_standard}$ ;  
se *cumulativo*=FALSO, dà il valore  $f(x)$  dove  $f$  è la densità di  $X$ .



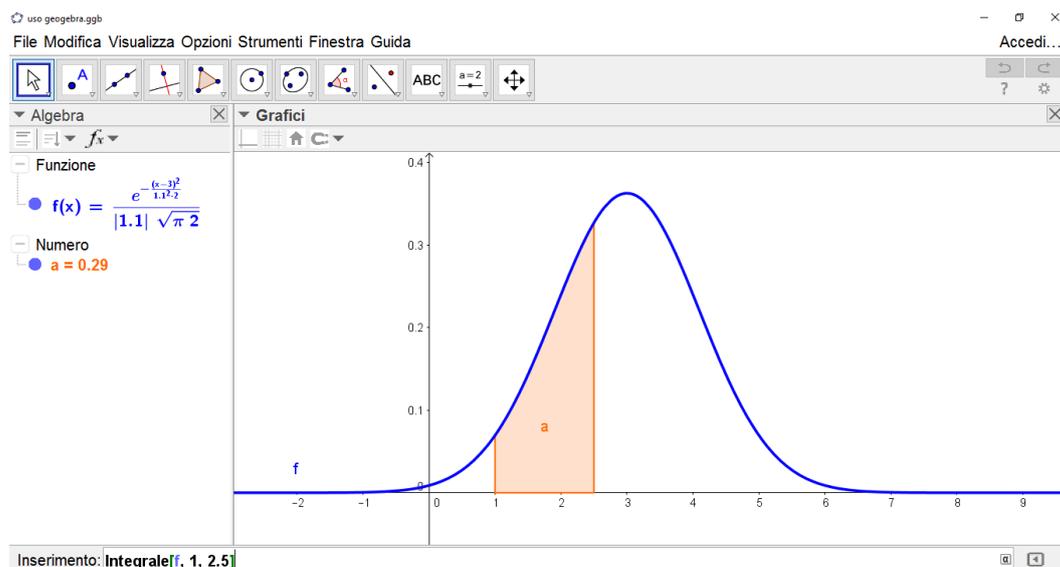
Nelle versioni che precedono Excel 2010, il nome è *DISTRIB.NORM.ST*.

Un ulteriore interessante strumento è **Geogebra**.

Questo software, infatti, permette di calcolare valori di probabilità (non solo relativi alla v.a. normale) come *integrali* della funzione densità.

Precisamente, inserita la funzione di densità (nella situazione in esame, una data densità normale<sup>25</sup>), si digita nella barra di inserimento il comando che serve, in generale, per calcolare l'integrale<sup>26</sup> di funzioni. L'aspetto interessante è che, oltre a restituire il risultato numerico, Geogebra visualizza il sottografico della densità nell'intervallo di integrazione.

*Tale strumento ti consente di controllare i procedimenti di calcolo relativi alla v.a. normale.*



<sup>25</sup>Comando *Normale*[<Media>, <Deviazione Standard>, <x>]. Questo comando, come gli altri, va digitato nella barra di inserimento. Una volta dato l'enter, viene visualizzato il grafico e l'espressione analitica della funzione che Geogebra indica ad esempio con  $f$ .

<sup>26</sup>Comando *Integrale*[ <Funzione>, <x iniziale>, <x finale>]; Geogebra indica con  $a$  (per esempio) il valore di tale integrale.

## 6 Applicazioni

### 6.1 Un problema diretto: la produzione di barre

Una macchina produce barre di acciaio a sezione circolare la cui lunghezza ottimale dovrebbe essere di 5 metri ed il diametro della sezione di 4 centimetri. Le barre effettivamente prodotte, che si suppongono tra loro indipendenti, hanno una lunghezza aleatoria con distribuzione normale di media  $\mu_1 = 5$  m e scarto standard  $\sigma_1 = 4$  cm. Il diametro della sezione è una variabile aleatoria, indipendente dalla precedente, e con distribuzione normale di media  $\mu_2 = 4$  cm e scarto standard  $\sigma_2 = 0,8$  cm. Una generica barra prodotta può essere direttamente venduta senza modifiche se la sua lunghezza è compresa tra 4,95 m e 5,05 m e la sua sezione tra 2,8 cm e 5,2 cm.

- a) Qual è la probabilità di poter vendere senza modifiche una generica barra prodotta?<sup>a</sup>
- b) Su 55.000 barre prodotte, quante ci si aspetta di poterne vendere senza modifiche?

<sup>a</sup>La questione è tratta dal problema n.3 della sessione ordinaria PNI, Esame di Stato 1998

- a) Sia  $L$  la v.a normale "lunghezza della barra" e sia  $D$  la v.a. normale "sezione della barra". La barra può essere venduta senza modifiche se (e solo se) vale:

$$4,95 \text{ cm} \leq L \leq 5,05 \text{ cm} \quad \text{e} \quad 2,8 \text{ cm} \leq D \leq 5,2 \text{ cm}$$

Il quesito richiede di calcolare la probabilità che si verifichino entrambe le condizioni. Valutiamo innanzitutto la probabilità

$$\mathbf{P(4,95 \leq L \leq 5,05)}$$

1. Standardizziamo la v.a  $L$

$$\begin{aligned} P(4,95 \leq L \leq 5,05) &= P\left(\frac{4,95 - 5}{0,04} \leq Z \leq \frac{5,05 - 5}{0,04}\right) \\ &= P(-1,25 \leq Z \leq 1,25) \end{aligned}$$

2. Per la simmetria del grafico della densità normale si ha

$$\begin{aligned} P(-1,25 \leq Z \leq 1,25) &= 2P(0 \leq Z \leq 1,25) \\ &= 2(P(Z \leq 1,25) - P(Z \leq 0)) \end{aligned}$$

3. E, ricorrendo alle tavole, otteniamo

$$2(P(Z \leq 1,25) - P(Z \leq 0)) \simeq 2(0,894 - 0,5) = 0,788$$

Poi, in modo esattamente analogo, otteniamo

$$\mathbf{P(2,8 \leq D \leq 5,2)} \simeq 0,866$$

Pertanto la probabilità di poter mettere in vendita la barra senza modifiche è<sup>27</sup>

$$\begin{aligned} P(4,95 \leq L \leq 5,05 \text{ e } 2,8 \leq D \leq 5,2) &= P(4,95 \leq L \leq 5,05) \cdot P(2,8 \leq D \leq 5,2) \\ &\simeq 0,788 \cdot 0,866 \simeq \boxed{0,68} \end{aligned}$$

b) Su 50.000 barre ci si aspetta di poterne vendere senza modifiche, circa

$$55.000 \cdot 0,68 = \boxed{37.400}$$

Infatti possiamo interpretare il valore di probabilità ottenuto nel punto a) come un valore a cui si approssima la frequenza relativa dell'evento in esame, su un "grande" numero di prove, con probabilità "grande" (forma intuitiva della Legge dei grandi numeri).

*Osservazione.* In alternativa, possiamo pensare di modellizzare la situazione mediante la distribuzione binomiale, caratterizzata da  $n = 55.000$  prove, ciascuna con probabilità  $p = 0,68$  di successo. Il valore atteso della corrispondente variabile aleatoria è proprio  $np$  e, per l'interpretazione statistica del valore atteso,  $np$  è "la" approssimazione richiesta.

---

<sup>27</sup>Assumiamo che la v.a.  $L$  e la v.a.  $D$  siano indipendenti. In tale ipotesi la legge della moltiplicazione afferma che l'intersezione dei due eventi in questione è il prodotto delle loro singole probabilità di realizzarsi.

## 6.2 Un problema inverso: il periodo di funzionamento di un prodotto

La vita media di un tipo di dispositivo prodotto da una certa ditta ha distribuzione normale con media 61 mesi e deviazione standard 6 mesi. La ditta intende vendere il dispositivo con una garanzia: se smette di funzionare entro  $T$  mesi, verrà sostituito con uno nuovo.

Se la ditta vuole contenere entro il 2% la probabilità di sostituzione del prodotto, quale periodo di garanzia  $T$  può proporre?

Indichiamo con  $X$  la variabile aleatoria che rappresenta la vita media del dispositivo. Per ipotesi essa ha distribuzione normale con parametri  $\mu = 61$  mesi e  $\sigma = 6$  mesi.

1. Operiamo prima sulla v.a. normale standard  $Z$ , in modo da poter ricorrere alle tavole della distribuzione di probabilità. In termini della v.a.  $Z$ , la richiesta del quesito si traduce nel chiedere *per quale valore  $k < 0$  di  $Z$  si ha*

$$P(Z \leq k) = 0,02$$

Però le tavole forniscono valori approssimati di probabilità del tipo  $P(Z \leq k)$  solo per  $k > 0$ . Pertanto:

- cercheremo sulle tavole una approssimazione del valore  $k^+$  tale che

$$P(Z \leq k^+) = 0,98$$

- il valore  $k < 0$  richiesto sarà

$$k = -k^+$$

Ora, per  $k^+ = 2,05$  le tavole forniscono il valore di probabilità 0,97982. Dunque  $k = -2,05$  e possiamo dire che

$$P(Z \leq -2,05) \simeq 0,02$$

2. Per concludere basta esprimere il risultato ottenuto in termini della v.a.  $X$ : al valore  $z = -2,05$  corrisponde il valore

$$x = \mu + \sigma z = 61 - 6 \cdot 2,05 = 48,7 \text{ mesi}$$

Ciò comporta che

$$P(X \leq 48,7) \simeq 0,02$$

Possiamo così concludere che il dispositivo smetterà di funzionare entro il tempo

$$\boxed{T = 48 \text{ mesi}}$$

con una probabilità minore del 2%. Perciò sono 48 i mesi di garanzia che la ditta può proporre stando nei limiti desiderati.

### 6.3 Altezze di una popolazione: aspetti quantitativi

Dall'elaborazione dei dati raccolti alla visita di leva dei nati nell'anno 1970, la statura risultò distribuita normalmente con media  $\mu \simeq 173,94$  cm e deviazione standard  $\sigma \simeq 8,63$  cm. Fino al 2014, per risultare abili all'arruolamento, era necessario avere una statura di almeno 165 cm.

Scelto a caso un giovane tra quelli appartenenti a tale classe di leva, qual è la probabilità che la sua statura fosse di almeno 165 cm?

Soluzione: la probabilità è circa 0,85

## 6.4 La curva normale: l'ordine nel caos -lettura-

Leggi il testo<sup>28</sup> di Mlodinov da pag.179 a pag.184 e schematizza gli aspetti matematici relativi alla curva normale che il testo mette in evidenza, soffermandoti sugli elementi nuovi che emergono. Ne riportiamo di seguito solo alcuni passi significativi<sup>29</sup>.

La distribuzione normale descrive il modo in cui molti fenomeni variano intorno a un valore centrale che ne rappresenta l'esito più probabile; nel suo *Essai philosophique sur les probabilités*, Laplace sosteneva che questa nuova matematica poteva essere usata per valutare l'attendibilità di testimonianze giuridiche, per prevedere i tassi di matrimonio e calcolare i premi assicurativi. Ma nell'ultima edizione di quell'opera, Laplace era già ultrasessantenne, quindi toccò a un uomo più

Studi statistici di questo tipo erano già stati condotti in precedenza, ma Quételet fece qualcosa di più con quei dati: non si limitò a calcolare le medie, ma analizzò il modo in cui i dati si discostavano dalla media. Ovunque guardasse, Quételet trovava la distribuzione normale: nella propensione al crimine, al matrimonio e al suicidio, nell'altezza degli indiani americani, nella circonferenza toracica dei soldati scozzesi (si imbatté in un campione di 5738 circonferenze toraciche in una vecchia copia dell'«Edinburgh Medical and Surgical Journal»). Nell'altezza di 100.000 giovani francesi chiamati alla leva trovò significativa una deviazione dalla distribuzione normale: se si tracciava un grafico per mettere in relazione il numero dei coscritti e la loro altezza, la curva a campana risultava distorta: ce n'erano troppo pochi appena sopra il metro e 56, e troppi appena al di sotto di quell'altezza. Quételet ipotizzò che la differenza – circa 2000 «uomini bassi» in più – fosse dovuta a frode, o magari a qualche favore concesso, dato che gli uomini sotto il metro e 56 venivano riformati.

<sup>28</sup>La lettura che proponiamo è tratta da *La passeggiata dell'ubriaco* di L. Mlodinov [34, pag.179-184].

<sup>29</sup>Ciò è dovuto a ragioni di spazio, ma è opportuno che il testo indicato venga letto per intero.

esempio meno noto è quello di Justin Wolfers, economista della Wharton School, che ha trovato indizi di frode sportiva nei risultati di circa 70.000 partite di basket del college.<sup>18</sup>

I point spread sono fissati dai bookmaker, ma in realtà sono determinati dalla massa degli scommettitori, perché i bookmaker li calcolano in modo da bilanciare la domanda. (I bookmaker guadagnano sulle commissioni, e cercano di far scommettere la stessa cifra su entrambe le squadre in modo da trarne un profitto comunque vada a finire l'incontro). Per misurare con che efficacia gli scommettitori valutano due squadre, gli economisti usano un numero chiamato «errore di previsione», che è la differenza tra il margine di vittoria della squadra favorita e il point spread determinato dal mercato. Forse non sorprenderà che l'errore di previsione, essendo una tipologia di errore, segua la distribuzione normale. Wolfers ha calcolato che la media è 0, il che significa che i point spread non tendono né a sovrastimare né a sottostimare le squadre, e la deviazione standard è 10,9 punti, il che vuol dire che circa 2/3 delle volte il point spread si trova entro 10,9 punti dal margine di vittoria. (In uno studio sugli incontri professionali di football si è scoperto un risultato simile, con una media di 0 e una deviazione standard di 13,9 punti.)<sup>19</sup>

Quando Wolfers ha esaminato il sottoinsieme di giochi che riguardavano squadre molto favorite, ha scoperto qualcosa di sorprendente: erano troppo poche le partite in cui la super-favorita vinceva con un margine poco superiore al point spread, ed erano davvero troppe le partite in cui la favorita vinceva di poco meno dello spread. Tornava in scena l'anomalia di Quételet. Anche Wolfers, come Quételet e Poincaré, è giunto alla conclusione che doveva trattarsi di frode. La sua

## 7 Il TLC

A questo punto del percorso, disponiamo di vari strumenti relativi alla v.a. normale. Essi ci consentiranno di precisare la questione di approssimazione introdotta nel paragrafo 1.3. In tale sezione abbiamo visto che la *distribuzione binomiale* si può approssimare con una opportuna *distribuzione normale*. Ma

*in che senso* la distribuzione binomiale è approssimata dalla distribuzione normale ?

Precisare tale approssimazione è un requisito indispensabile per effettuare concretamente l'approssimazione. In particolare relativamente al problema dei sondaggi che ci siamo posti all'inizio e che costituisce il filo conduttore.

### 7.1 Esplorazioni: i parametri -attività-

La nostra investigazione sull'approssimazione della binomiale inizia dall'esame dei valori da attribuire ai parametri della variabile aleatoria normale.

Precisamente, consideriamo una v.a. binomiale di parametri  $n$  e  $p$  fissati, dove  $n$  rappresenta il numero di prove e  $p$  la probabilità di successo nella prova.

*Per quali valori dei parametri  $\mu$  e  $\sigma^2$  della v.a. normale, la densità approssima la distribuzione binomiale?*

La forma dei due grafici suggerisce che le medie (normale e binomiale) debbano coincidere. Ossia ci aspettiamo valga

$$\mu = np$$

Resta dunque da individuare un valore per il parametro  $\sigma^2$ . È ciò che ci proponiamo di fare con l'attività seguente.

**Nota.** Per semplicità diciamo di confrontare il grafico della distribuzione binomiale con quello della densità normale. Ma distribuzione di probabilità e densità sono oggetti matematici diversi. Preciseremo più avanti in che senso vada considerato tale confronto.

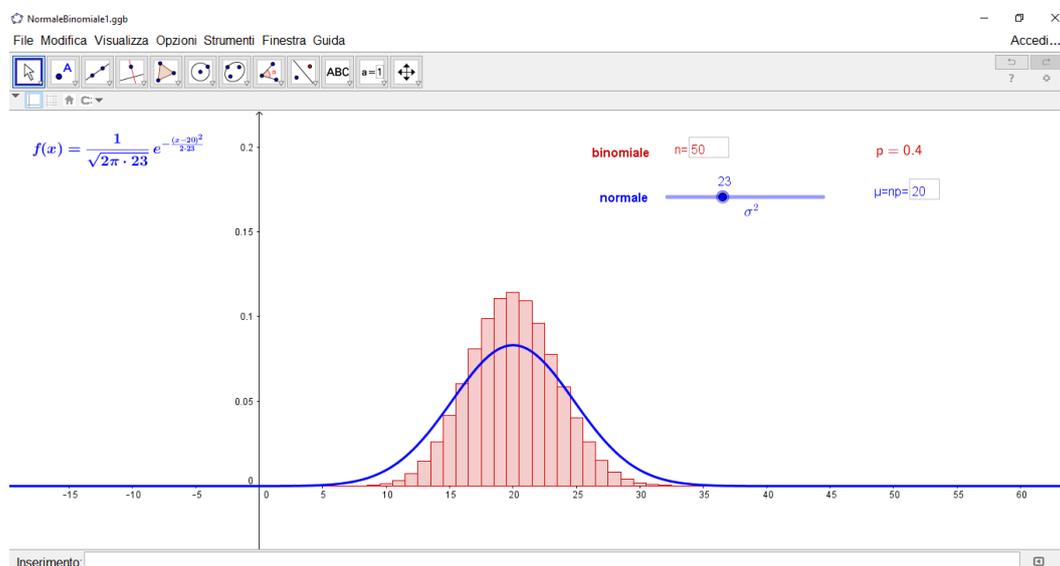
Utilizza il file Geogebra [NormaleBinomiale1.ggb](#) per rispondere alle seguenti questioni.

Considera le v.a. binomiali di media fissata  $np = 20$  e per i seguenti valori di  $n$ :  $n = 25$ ,  $n = 50$ ,  $n = 100$  e  $n = 200$ .

Considera poi la v.a. normale che ha media  $\mu = np = 20$ .

- In ciascuno dei casi indicati, determina un valore di  $\sigma^2$  per il quale la curva normale sembra approssimare la distribuzione binomiale.
- Quale relazione ti sembra possa esserci tra il valore trovato di  $\sigma^2$  e quello dei parametri  $n$ ,  $p$  della binomiale?

### Come si utilizza il file NormaleBinomiale1.ggb



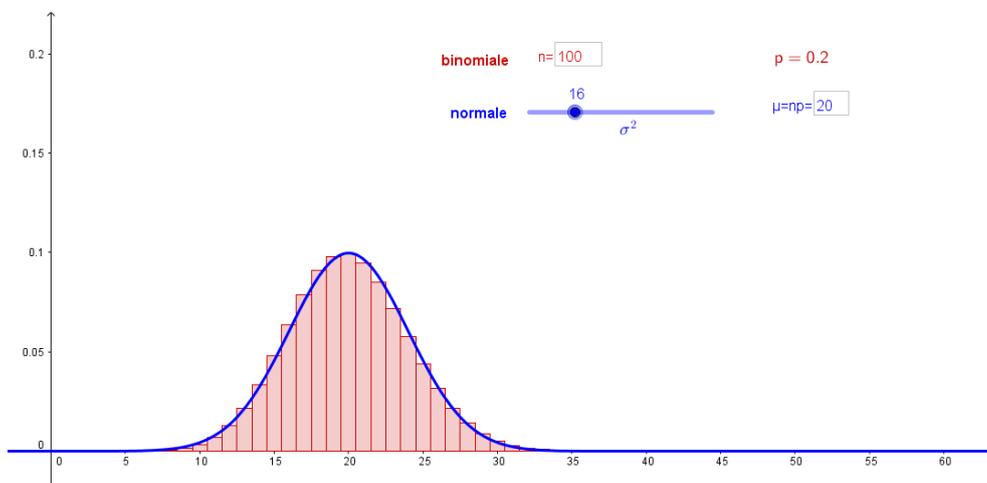
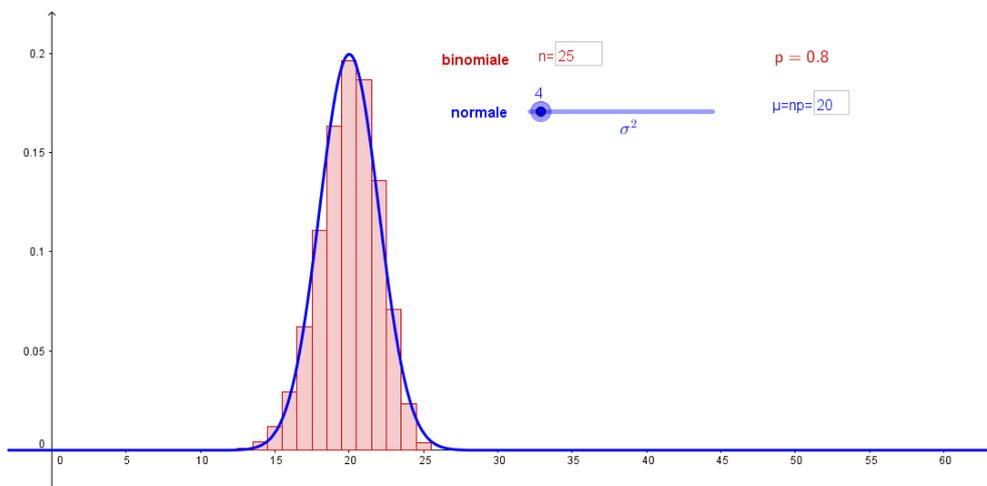
Il file dispone di:

- una casella di inserimento per il parametro  $n$  della **binomiale**: a seconda del valore di  $n$  qui inserito, ottieni (automaticamente) il grafico della distribuzione binomiale;
- uno slider per il parametro  $\sigma^2$  della **normale**: al variare di  $\sigma^2$  (muovi il cursore) si ottengono i corrispondenti grafici della densità normale.

## Risoluzione

a) Esaminando i grafici, si dovrebbe osservare che si ha una "buona" approssimazione intorno a:

- $\sigma^2 = 4$  quando  $n = 25$
- $\sigma^2 = 12$  quando  $n = 50$
- $\sigma^2 = 16$  quando  $n = 100$
- $\sigma^2 = 18$  quando  $n = 200$



- b) Facendo un po' di conti, pare che il valore di  $\sigma^2$ , individuato nel punto a) nei vari casi, sia proprio uguale a  $np(1 - p)$ . Ma quest'ultima espressione rappresenta la varianza della v.a. binomiale. Pertanto, almeno nei casi esaminati, sembra che si abbia una "buona" approssimazione quando le varianze delle due v.a. coincidono.

## Conclusioni

Il risultato congetturato nell'attività precedente vale più in generale.

Se la varianza della v.a. normale è uguale a quella della v.a. binomiale (e le medie delle due v.a. coincidono), si ha una "buona" approssimazione della distribuzione binomiale. Ossia per

$$\sigma^2 = np(1 - p)$$

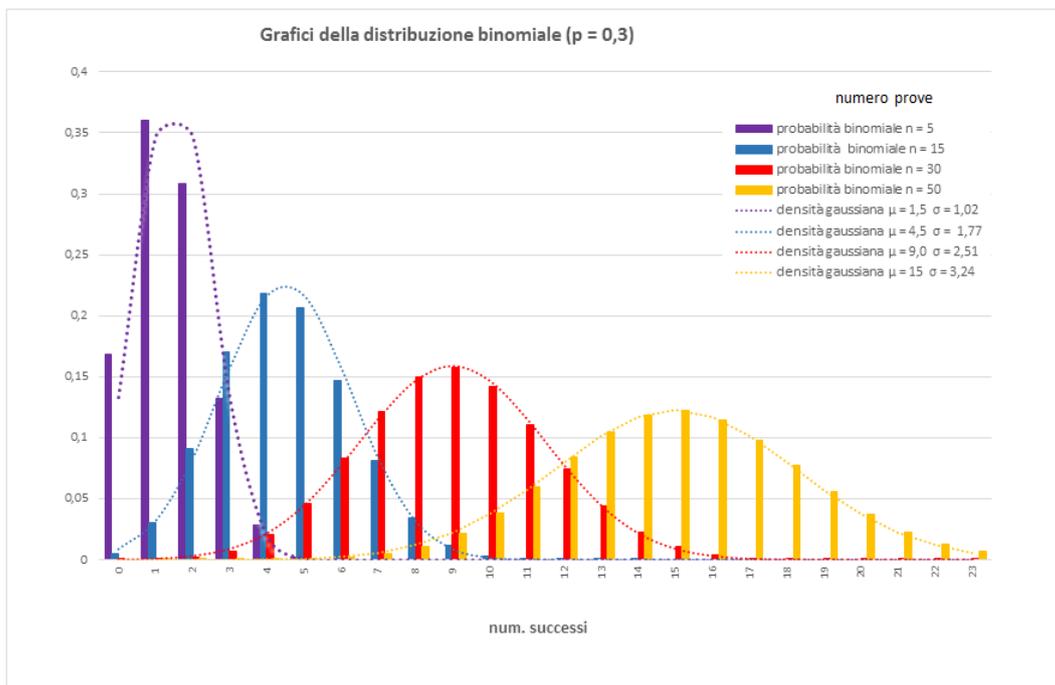
Attenzione però: come vedremo nella prossima attività, l'uguaglianza dei parametri della binomiale e della normale *non è sufficiente* per avere una "buona" approssimazione.

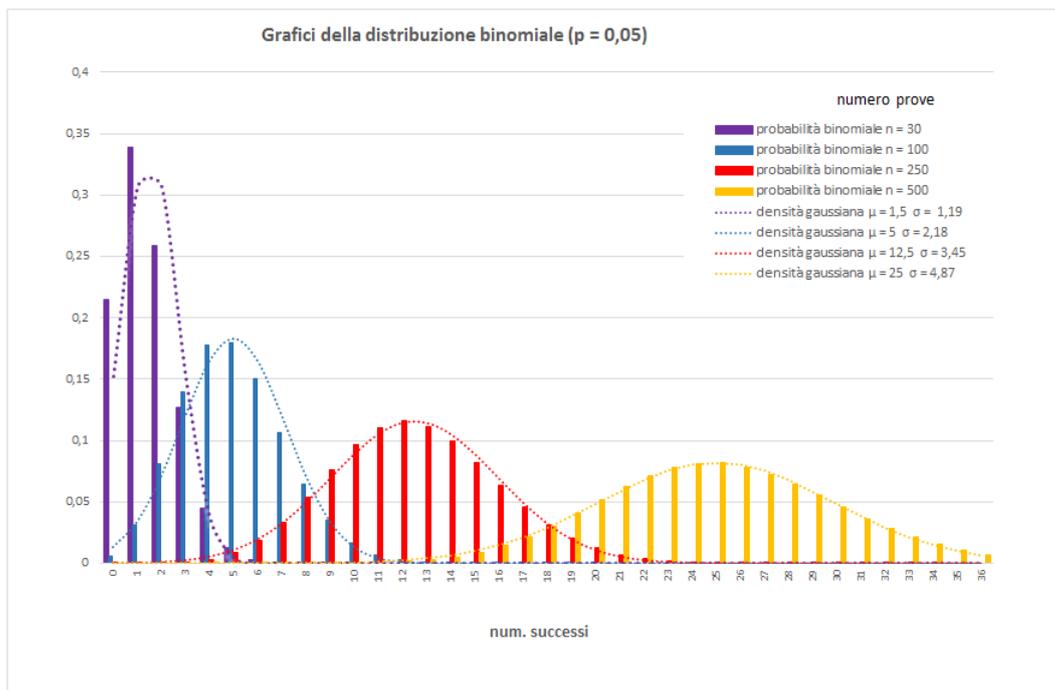
## 7.2 Esplorazioni: il numero di prove -attività-

*Una volta stabilito il legame tra i parametri  $\mu$  e  $\sigma^2$  della v.a. normale e i parametri  $n$  e  $p$  della v.a. binomiale, sondiamo come varia l'approssimazione al variare del numero di prove.*

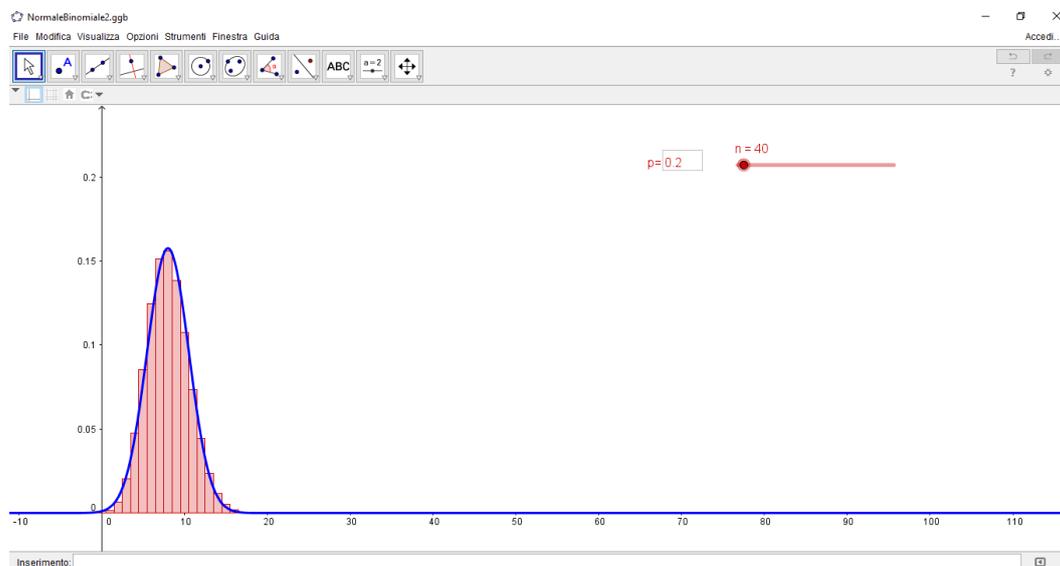
Nelle seguenti figure fissiamo prima  $p = 0,3$  e poi  $p = 0,05$ . In entrambi i casi confrontiamo i grafici (realizzati con Excel) relativi alla v.a. binomiale e alla v.a. normale, per valori crescenti di  $n$ .

Osserviamo che con  $p = 0,05$  occorre un  $n$  più grande, rispetto al caso  $p = 0,3$ , perché la distribuzione binomiale assomigli ad una curva "a campana".





Usa il file Geogebra [NormaleBinomiale2.ggb](#) per effettuare alcune prove.



I grafici esaminati suggeriscono che l'approssimazione è "buona" per  $n$  "grande".

## Conclusioni

Ciò vale in generale, cioè

per  $n$  "grande", è "buona" l'approssimazione della distribuzione binomiale con la normale (che ha la stessa media e varianza della binomiale).

### 7.3 La sostanza del teorema

Le due conclusioni a cui siamo giunti negli ultimi due paragrafi si possono sintetizzare nel seguente teorema.

#### TEOREMA LIMITE CENTRALE -TLC (formulazione semplificata)

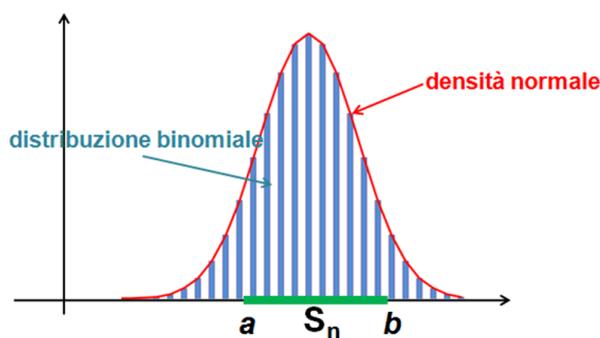
Sia  $S_n$  la v.a. binomiale di parametri  $n$  e  $p$ , dove  $n$  rappresenta il numero di prove e  $p$  la probabilità di successo nella prova.

Sia  $X$  la v.a. normale che ha media  $\mu$  e varianza  $\sigma^2$  uguali rispettivamente alla media e alla varianza di  $S_n$ . Ossia

$$\mu = np \quad \text{e} \quad \sigma^2 = np(1 - p)$$

Allora per  $n$  "grande", vale l'approssimazione:

$$\mathbf{P(a \leq S_n \leq b)} \simeq \mathbf{P(a \leq X \leq b)}$$



**Nota.** In figura stiamo confrontando una *distribuzione di probabilità* (quella binomiale) con una *densità di probabilità* (quella normale). La figura suggerisce che i loro valori siano "vicini", ma esse sono oggetti matematici di tipo diverso, mentre nell'enunciato del TLC intervengono solo valori di probabilità. Pertanto, a rigore, dovremmo interpretare la distribuzione binomiale come una funzione costante a tratti sull'intervallo reale  $[a; b]$ : con questo accorgimento le due probabilità confrontate nel TLC si possono leggere in figura come aree dei corrispondenti sottografici in  $[a; b]$ .

## 7.4 Applicazioni: finalmente i sondaggi

Torniamo al problema guida proposto nel paragrafo 1.1.

Una popolazione costituita da **10.000** individui è chiamata a votare tra due candidati, diciamo  $A$  e  $B$ . Mediante un sondaggio effettuato su un campione della popolazione, si è stimato che la probabilità che l'**individuo** sia favorevole ad  $A$  è del **40%**.

Sulla base di ciò, si vuole stimare il numero **F** di individui che voteranno  $A$ . Precisamente, quanto vale

$$\mathbf{P(3900 \leq F \leq 4100) ?}$$

Ora mediante il TLC possiamo risolverlo in modo computazionalmente meno articolato:

1. modellizziamo la situazione mediante la distribuzione binomiale, come già fatto all'inizio del percorso; essa ha parametri  $n = 10.000$  e  $p = 0,4$
2. per il TLC, il numero di individui  $F$  si approssima mediante la v.a. normale  $X$  di parametri

$$\mu = np = 4.000 \quad \text{e} \quad \sigma^2 = np(1 - p) = 2400$$

3. calcoliamo quindi la probabilità richiesta

$$\begin{aligned} \mathbf{P(3900 \leq F \leq 4100)} &\stackrel{\substack{\approx \\ \uparrow \\ \text{per il TLC}}}{\approx} P(3900 \leq X \leq 4100) \stackrel{\substack{= \\ \uparrow \\ \text{standardizzando } X}}{=} P\left(\frac{3900-4000}{\sqrt{2400}} \leq Z \leq \frac{4100-4000}{\sqrt{2400}}\right) \\ &\simeq P(-2,04 \leq Z \leq 2,04) \stackrel{\substack{\approx \\ \uparrow \\ \text{ricorrendo alle tavole}}}{\approx} 0,9586 \end{aligned}$$

Pertanto<sup>30</sup> la probabilità che la percentuale di cittadini favorevoli ad  $A$  sia compresa tra il 39% e il 41% dell'intera popolazione è circa  $\boxed{0,9586}$ .

<sup>30</sup>Nell'ultimo passaggio, conviene calcolare  $P(-2,04 \leq Z \leq 2,04)$  sfruttando la simmetria della densità normale e dell'intervallo di variabilità di  $Z$ . Pertanto tale probabilità è uguale a  $2P(0 \leq Z \leq 2,04) = 2(P(Z \leq 2,04) - P(Z \leq 0))$ . Dalle tavole si deduce che  $P(Z \leq 2,04) \simeq 0,9793$ ; invece per  $P(Z \leq 0)$  **non serve** ricorrere alle tavole.

### Osservazione: i due approcci a confronto

- Il TLC permette di risolvere il problema iniziale senza dover svolgere molti conti. Invece il ricorso al modello binomiale, come si accorse De Moivre, comporta parecchie valutazioni di probabilità ( $P(F = 3900) + P(F = 3901) + \dots + P(F = 4100)$ ).
- Proviamo comunque a ricorrere alla sola distribuzione binomiale e usiamo un calcolatore<sup>31</sup> per portare a termine il calcolo impostato già nel paragrafo 1.1. Risulta

$$P(3900 \leq F \leq 4100) \simeq 0,95877$$

I due numeri (quello trovato con la normale e quello trovato con la binomiale) sono "vicini" ma comunque diversi. Qual è quello "giusto"?

In realtà ... nessuno dei due. Infatti i due valori di probabilità sono stati ottenuti all'interno di due modelli diversi, binomiale e normale, che sono due *schematizzazioni* della situazione reale, tra le varie *possibili*. E tali modelli sono stati costruiti sulla base di nostre precise *scelte* sulla situazione reale: non sono la situazione reale. Concludiamo così che nessuno dei due valori è "più giusto" dell'altro e che entrambi rappresentano due possibili soluzioni del problema in esame.

---

<sup>31</sup>Con Excel si può usare la funzione `DISTRIB.BINOM.N(num_successi;prove;probabilità_s;cumulativo)` con `cumulativo=VERO`.

## 7.5 Applicazioni: dall'esame di Stato

Il TLC permette di risolvere in modo più agevole anche il quesito n.3 proposto nella simulazione della seconda prova<sup>32</sup> il 29 aprile 2016.

Durante il picco massimo di un'epidemia di influenza, il 15% della popolazione è a casa ammalato:

- a) qual è la probabilità che in una classe di 20 alunni ce ne siano più di due assenti per influenza?
- b) descrivi le operazioni da compiere per verificare che, se l'intera scuola ha 500 alunni, la probabilità che ce ne siano più di 50 influenzati è maggiore del 99%.

- a) *Decidiamo*<sup>33</sup> di modellizzare la situazione mediante la distribuzione binomiale. Consideriamo dunque la variabile aleatoria binomiale

$$S = \text{numero di assenti per malattia}$$

I valori dei parametri della distribuzione sono:

$$\text{numero di prove } \mathbf{n = 20}$$

$$\text{probabilità che l'alunno si ammali } \mathbf{p = 0,15}.$$

Secondo il modello così costruito la probabilità dell'evento "vi sono più di 2 assenti per malattia" è<sup>34</sup>

$$\begin{aligned} P(S > 2) &= 1 - P(S \leq 2) \\ &= 1 - P(S = 2) - P(S = 1) - P(S = 0) \\ &= 1 - (190 \cdot 0,15^2 \cdot 0,85^{18} + 20 \cdot 0,15 \cdot 0,85^{19} + 0,85^{20}) \\ &\simeq \boxed{0,5951} \end{aligned}$$

- b) Proviamo prima a schematizzare la situazione mediante un modello binomiale analogo al precedente. In questo caso i valori dei parametri della distribuzione sono:

$$\text{numero di prove } \mathbf{n = 500}$$

$$\text{probabilità che l'alunno si ammali } \mathbf{p = 0,15}.$$

Secondo tale modello la probabilità dell'evento "vi sono più di 50 ammalati" è:

$$\begin{aligned} P(S > 50) &= 1 - P(S \leq 50) \\ &= 1 - P(S = 50) - P(S = 49) - \dots - P(S = 0) \end{aligned}$$

Il calcolo è dunque articolato.

<sup>32</sup>Tale quesito era già stato assegnato nella sessione straordinaria italiana di settembre 2015 e nella sessione ordinaria 2015, scuole italiane all'estero, Americhe.

<sup>33</sup>In effetti di decisione si tratta: sulla base delle informazioni fornite nel testo del quesito si possono costruire diversi **modelli**; la distribuzione binomiale è solo **uno** dei modelli possibili.

<sup>34</sup>Per la prima uguaglianza siamo passati all'evento complementare.

Per semplificarlo ricorriamo al TLC<sup>35</sup> e approssimiamo la distribuzione binomiale con la distribuzione normale che ha la stessa media  $\mu$  e la stessa varianza  $\sigma^2$  della binomiale considerata. Ossia

$$\mu = np = 75 \quad \text{e} \quad \sigma^2 = np(1-p) = 63,75$$

Detta  $X$  la variabile aleatoria normale "*numero di ammalati*", vale

$$\begin{array}{ccccccc}
 \mathbf{P(S > 50)} & \simeq & P(X > 50) & = & P(Z > \frac{50-75}{\sqrt{63,75}}) & \simeq & P(Z > -3,13) \\
 & \uparrow & & \uparrow & & & \\
 & \text{per il TLC} & & \text{standardizzando } X & & & \\
 & & & & & & \\
 & & = & P(Z < 3,13) & \simeq & \boxed{0,9991} \\
 & \uparrow & & \uparrow & & & \\
 & \text{per simmetria} & & \text{ricorrendo alle} & & & \\
 & & & \text{tavole} & & & 
 \end{array}$$

Abbiamo così verificato che la probabilità che ci siano più di 50 influenzati su una scuola di 500 alunni è maggiore del 99%.

---

<sup>35</sup>In realtà si poteva applicare il TLC anche nella situazione descritta nel punto a) del quesito. In tal caso, però, i vantaggi computazionali che comporta tale approccio non sono così evidenti come nella situazione prospettata nel punto b).

## Riferimenti bibliografici

### Riferimenti relativi alla didattica

#### Siti

- [1] *Indicazioni nazionali per il secondo ciclo di istruzione*, 2010.  
[http://archivio.pubblica.istruzione.it/riforma\\_superiori/nuovesuperiori/index.html](http://archivio.pubblica.istruzione.it/riforma_superiori/nuovesuperiori/index.html)
- [2] *Linee guida per l'elaborazione dei piani di studio per i Licei*, Provincia autonoma di Trento, 2013.  
<https://www.vivoscuola.it/lineeguida-secondociclo-licei>
- [3] UMI-CIIM, *Proposta di un Syllabus di matematica per i Licei Scientifici (nuovo ordinamento)*, 2014.  
<http://www.umi-ciim.it/materiali-umi-ciim/secondo-ciclo>
- [4] *Corso di formazione in didattica della probabilità - DiCoMat Lab in collaborazione con IPRASE*.  
<http://www.iprase.tn.it/formazione/formazione-docenti-e-dirigenti/corsi/didattica-della-probabilita-per-la-secondaria-di-secondo-grado/>
- [5] *Laboratorio di Didattica e Comunicazione della Matematica*, Università degli Studi di Trento.  
<http://r.unitn.it/it/maths/dicomatlab>
- [6] *Moocs for teacher series: "To Flip Or Not To Flip - Discover the flipped classroom methodology"*, Politecnico di Milano.  
[https://www.pok.polimi.it/courses/course-v1:Polimi+FC101+2016\\_M9/about](https://www.pok.polimi.it/courses/course-v1:Polimi+FC101+2016_M9/about)
- [7] BARRA M. *Teorema del limite centrale a scuola a partire dall'esperienza e con il problem solving. Somma di alcuni numeri aleatori e Metodo di Montecarlo*.  
<http://www.dmmm.uniroma1.it/accascinamonti/ssis/linguaggiodelincertezza2/1%20TeorLimiteCentrStampa2.pdf>
- [8] BOLONDI G., VIVIANI *Dispense da corso formazione: aspetti linguistici della matematica*  
[https://www.google.it/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwid2MCMjrvKAhVGmw4KHc-2CrIQFggfMAA&url=http%3A%2F%2Fwww.iprase.tn.it%2Falfresco%2Fgd%2Fa%2Fworkspace%2FspacesStore%2F0bcc9cab-bbff-4557-b764-f90eb42d188f%2Fmateriali%2520Viale%2520Matteo.pdf&usg=AFQjCNEO\\_RsuVfrT5n\\_zg4tV3UirPhpsPg&bvm=bv.112064104,d.ZWU](https://www.google.it/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwid2MCMjrvKAhVGmw4KHc-2CrIQFggfMAA&url=http%3A%2F%2Fwww.iprase.tn.it%2Falfresco%2Fgd%2Fa%2Fworkspace%2FspacesStore%2F0bcc9cab-bbff-4557-b764-f90eb42d188f%2Fmateriali%2520Viale%2520Matteo.pdf&usg=AFQjCNEO_RsuVfrT5n_zg4tV3UirPhpsPg&bvm=bv.112064104,d.ZWU)

#### Testi

- [9] ANZELLOTTI G. *Valutazione e sviluppo delle competenze matematiche di base dall'obbligo scolastico all'ingresso dell'università*, in *RICERCAZIONE Ricerca educativa, valutativa e studi sociali sulle politiche e il mondo giovanile*, Edizioni Erickson, v.3, n.1 Giugno 2011, pp.173-184.

- [10] ANZELLOTTI G., CAPPELLO L., INNOCENTI S. *Matematica: obiettivi, itinerari, interpretazioni*, in *Nuova secondaria*, Editrice La Scuola, v.23, n.1 (2005), pp.91-97.
- [11] BARRA M. *Parliamo di probabilità e del suo insegnamento*, Progetto Alice, v.XVII, n.49, 2016.
- [12] D'AMORE B. *Elementi di Didattica della Matematica*, Pitagora Editrice Bologna, 1999.
- [13] D'AMORE B., SBARAGLI S. *Analisi semantica e didattica dell'idea di "misconcezione"* in *La Matematica e la sua Didattica*, n.2/2005, pp. 139-163.
- [14] FREUDENTHAL H. *Ripensando l'educazione matematica*, Edizioni La Scuola, 1994.
- [15] LANG S. *La bellezza della matematica*, Bollati Boringhieri, 1991.
- [16] LOLLI G. *Se viceversa*, edizioni Bollati Boringhieri, 2014
- [17] MARIOTTI M.A., MAFFEI L. *Difficoltà in algebra: un intervento di recupero, Parte Prima* in *La Matematica e la sua Didattica*, n.1/2006, pp. 81-99.
- [18] MORIN E. *Insegnare a vivere: manifesto per cambiare l'educazione*, Raffaello Cortina Editore, 2015.
- [19] PELLERREY M. *Le competenze cosa sono* in *L'insegnamento della matematica e delle scienze integrate*, Vol. 38 n.5, 2015
- [20] ROSSI C. *La matematica dell'incertezza. Didattica della probabilità e della statistica*, Zanichelli, 1999.
- [21] SFARD A. *Psicologia del pensiero matematico*, Edizioni Erickson, 2009.
- [22] VILLANI V., BERNARDI C., ZOCCANTE S., PORCARO R. *Non solo calcoli. Domande e risposte sui perché della matematica*, Edizioni Springer, 2012.
- [23] ZAN R. *Difficoltà in matematica. Osservare, interpretare, intervenire*, Edizioni Springer, 2007.
- [24] ZAN R. *I danni del "bravo" insegnante*, in *Le difficoltà in Matematica: da problema per pochi a risorsa per tutti*, Pitagora Editrice, Bologna, 2001.
- [25] ZANETTI V. *La fisica attorno a noi*, Zanichelli, Bologna, 1989.

### **Libri di testo per la scuola secondaria di secondo grado**

- [26] BARONCINI P., MANFREDI R., FRAGNI I. *Lineamenti.Math.blu*, Volume 5, Ghisetti e Corvi, 2012.
- [27] BERGAMINI M., TRIFONE A., BAROZZI G. *Matematica.blu 2.0*, Zanichelli, 2011.
- [28] MARASCHINI W., PALMA M. *multi ForMat*, Modulo 24, Paravia, 2002.
- [29] SASSO L. *La matematica a colori*, Edizione blu per il quinto anno, Petrini, 2015.

## Riferimenti relativi al calcolo delle probabilità

- [30] BALDI P. *Introduzione alla probabilità*, McGraw-Hill, Seconda Edizione, 2012.
- [31] BONACCORSI S. *Appunti di probabilità*, Dipartimento di Matematica, Università degli Studi di Trento, anno accademico 2004/05.
- [32] CICCHITELLI G. *Probabilità e statistica*, Maggioli, Prima Edizione, 1990.
- [33] GRINSTEAD C.M., SNELL J.L. *Introduction to probability*, AMS, 1997.
- [34] MLODINOW L. *La passeggiata dell'ubriaco. Le leggi scientifiche del caso*, Rizzoli, 2009.
- [35] PIAZZA R. *I capricci del caso*, Edizioni Springer, 2009.
- [36] PRODI G. *Metodi matematici e statistici*, McGraw-Hill, 1992.
- [37] TALEB N. N. *Il cigno nero*, ilSaggiatore, 2007.