

# Il teorema limite centrale

I limiti in probabilità

- 0. Introduzione
  - 1. La distribuzione binomiale
  - 2. La distribuzione gaussiana e il teorema limite centrale
  - 3. Dalla media statistica alla media teorica
  - 4. La distribuzione esponenziale
  - 5. Approfondimenti
  - 6. Esercizi
- ➔ Sintesi

## 0. Introduzione

In questa scheda approfondiremo lo studio di alcune variabili casuali discrete e continue (concetti introdotti nella scheda per la classe 2ª ➔ *Calcolo delle probabilità*) che abbiamo avviato nella ➔ scheda *Quale matematica per i fenomeni casuali?* Vedremo come, in molti casi, vi siano variabili casuali che possono essere interpretate come la somma di altri tipi di variabili casuali e che, quando queste sono molte, tendono a distribuirsi secondo un istogramma dalla forma a campana, che può essere approssimato dal grafico di una opportuna funzione.

## 1. La distribuzione binomiale

Consideriamo il problema:

*un forno automatico produce mediamente 1 biscotto bruciato ogni 8, e i biscotti sono successivamente mescolati e impacchettati automaticamente in confezioni da 6,*

- qual è la probabilità che in una confezione non vi siano biscotti bruciati, o che ve ne sia 1, ..., o che siano tutti bruciati?,
- cioè qual è la legge di distribuzione della variabile casuale  $N = \text{"n° di biscotti bruciati in una confezione"}$ ?

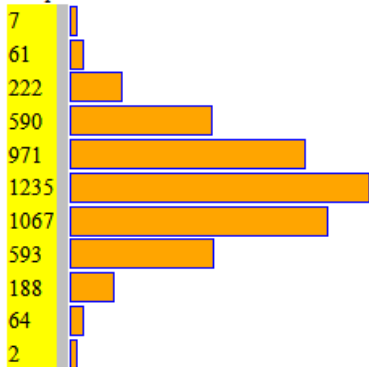
È un problema inventato, ma che dà bene l'idea di molti problemi pratici che in una attività produttiva occorre affrontare; in questa sezione metteremo a punto gli strumenti, semplici, per farlo. Partiamo, come in genere si fa, da un problema più semplice.

Ecco una situazione abbastanza simile a quella delle successive alzate di un mazzo di carte:

*lanciamo una moneta equa 10 volte; qual è la legge di distribuzione della variabile casuale:  $N = \text{"n° di teste uscite"}$ ?*

Proviamo prima a studiare il problema con una **simulazione**. L'idea è:

A Freq.



- generare 10 volte un numero a caso equamente distribuito tra 0 e 1, pensando 0 come "croce" e 1 come "testa" e
- prendere come N la somma di questi 10 numeri; infatti tale somma è uguale a numero delle volte che è uscito 1, cioè "testa".

Ecco qua gli esiti e l'analisi di una tale simulazione di 5 lanci effettuata con lo script a cui puoi da ➔ qui, copiando poi gli esiti e incollandoli nello script [Istogramma](#)

A = -0.5 B = 10.5 intervals = 11 their width = 1 n=5000 min=0 max=10 median = 5  
1^|3^ quartile = 4|6 mean = 4.9974

B 5000

Effettuiamo, ora, lo **studio teorico**. Dobbiamo trovare i valori di  $\Pr(N=0)$ ,  $\Pr(N=1)$ , ...,  $\Pr(N=10)$ . Iniziamo a considerare, per es.,  $\Pr(N=5)$ .

- Valutiamo la probabilità che testa esca *esattamente nei primi 5 lanci* (vedi sotto a sinistra). La probabilità che esca T in un lancio è  $1/2$ , la stessa che esca C. Poiché i dieci lanci sono *indipendenti*, posso *moltiplicare le probabilità*. Ottengo che la probabilità cercata è  $(1/2) \cdot (1/2) \cdot \dots \cdot (1/2) = (1/2)^{10}$ , cioè  $2^{-10}$ .

T T T T T C C C C C || C T T T T C C C C  
.5 .5 .5 .5 .5 .5 .5 .5 .5 .5 || .5 .5 .5 .5 .5 .5 .5 .5 .5

- Ottengo lo stesso valore per la probabilità che testa esca *nei lanci tra il 2° e il 6°* (vedi sopra a destra), e per qualunque altra collocazione dei 5 lanci in cui esce testa.

- I modi in cui posso scegliere i 5 posti sono  $C(10,5)$ , dove  $C(n,k)$  indica quante sono le **combinazioni** di  $n$  elementi  $k$  a  $k$ , cioè la *quantità dei sottoinsiemi di  $k$  elementi di un insieme di  $n$  elementi* (vedi la ➔ scheda *La matematica tra gioco e realtà*).

- $C(10,5) = 10/5 \cdot 9/4 \cdot 8/3 \cdot 7/2 \cdot 6/1 = 9 \cdot 4 \cdot 7$

- Quindi, la probabilità cercata, per la proprietà additiva, è la somma di  $C(10,5)$  termini uguali a  $2^{-10}$

$$\Pr(N=5) = C(10,5) \cdot 2^{-10} = 9 \cdot 4 \cdot 7 / 2^{10} = 24.6\%$$

$$\text{In generale: } \Pr(N=i) = C(10,i) / 2^{10} \quad (i = 0, 1, \dots, 10)$$

Posso confrontare questo valore con quello ottenuto sperimentalmente: 1235 / 5000 \* 100 = 24.7 OK!

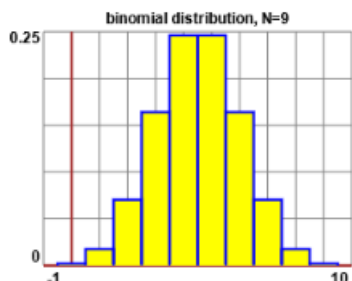
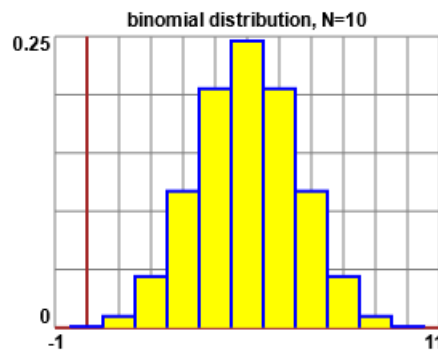
- 1 Confronta analogamente il valore di  $\Pr(N=3)$  dell'istogramma ottenuto con la simulazione con quello che ottieni con la formula precedente.

Se lanciassi  $n$  volte la moneta avrei, del tutto analogamente:

$$\Pr(N = i) = C(n, i) / 2^n \quad (i = 0, 1, \dots, n)$$

È un caso particolare di *legge di distribuzione binomiale*, discussa più in generale tra poco. Il nome deriva dalla presenza del coefficiente binomiale  $C(n, i)$ . Ecco, a destra, il grafico relativo al caso  $n = 10$ , lo stesso studiato sperimentalmente..

Il grafico è stato ottenuto con lo script [binom10](#) (vedi i commenti nella riga seguente).



Le altezze delle colonne (vedi i valori di  $P$  in [binom10](#)) sono stati ottenute con la [grande CT](#) calcolando i valori di  $C(10, 0), C(10, 1), \dots [1, 10, 45, 120, 210, 252, 210, \dots, 1]$ , di  $2^{10} [1024]$ , di  $1/1024 [0.0009765625]$ , e infine operando su  $1, 10, 45, 120, 210, 252, 210, 120, 45, 10, 1$  con `[data·p]` avendo messo  $0.0009765625$  in  $p$ .

2 Con [binom9](#) ottieni l'istogramma relativo al caso  $n = 9$ . Come sono stati calcolati i valori di  $P$  in questo caso?

Se metto in input  $0*1, 1*10, 2*45, 3*120, 4*210, 5*252, 6*210, 7*120, 8*45, 9*10, 10*1$  o  $0*1, 1*9, 2*36, 3*84, 4*126, 5*126, 6*84, 7*36, 8*9, 9*1$ , con la [grande CT](#) posso calcolare la *varianza* ottenendo nei due casi **2.5** e **2.25**. La media nei due casi vale **5** e **4.5**.

In tutti i casi, non solo per  $n=10$  e  $n=9$ , la *media*, per simmetria, è evidentemente  $n/2$ . È naturale congetturare che in tutti i casi la *varianza* sia  $n/4$  e che, quindi, lo *sclarto quadratico medio*, ossia la radice quadrata della varianza, sia  $\text{sqm} = \sqrt{n/2}$ . La cosa è dimostrata in §5.

Torniamo all'esempio del  $\rightarrow$  forno. Ragioniamo in modo simile a quanto fatto nel caso delle monete. Sia  $N$  la variabile casuale a valori in  $\{0, 1, \dots, 6\}$  che rappresenta il numero di biscotti difettosi. È sensato ritenere (dato il rimescolamento presente prima del confezionamento) che le estrazioni dei biscotti siano una indipendente dall'altra. Quindi la probabilità che esattamente i primi  $k$  biscotti estratti siano difettosi è data dal prodotto:

$$(1/8) \cdot (1/8) \cdot \dots \cdot (1/8) \cdot (7/8) \cdot \dots \cdot (7/8) = (1/8)^k \cdot (7/8)^{6-k}.$$

Poiché non ci interessa la disposizione dei biscotti difettosi, abbiamo:

$$\Pr(N=k) = C(6, k) \cdot (1/8)^k \cdot (7/8)^{6-k}.$$

Generalizzando dal caso di una confezione da 6 ad una da  $n$  biscotti, dal caso della difettosità con probabilità di  $1/8$  a quello della difettosità con probabilità  $p$ , abbiamo che la probabilità che in una confezione vi siano  $k$  biscotti difettosi è:

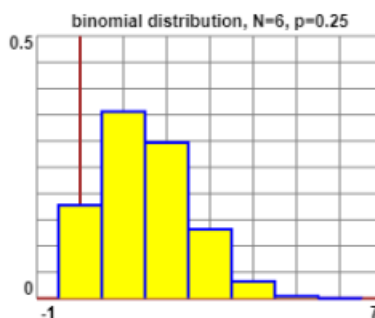
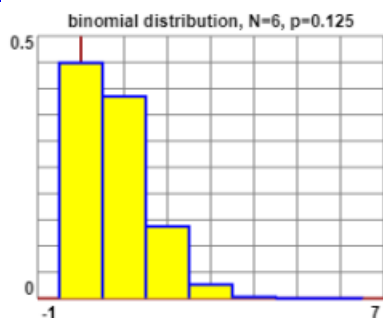
$$\Pr(N = k) = C(n, k) \cdot p^k \cdot (1 - p)^{n-k}$$

Anche questa legge di distribuzione, che generalizza quella considerata nel caso delle monete eque, viene chiamata *legge di distribuzione binomiale* (o di *Bernoulli*):

*si applica a tutte le situazioni in cui si ripete  $n$  volte la prova su una variabile casuale che può assumere solo due valori, in cui  $p$  è la probabilità di uno di questi due valori e  $N$  è il numero delle volte in cui questo valore esce.*

Ecco le elaborazioni grafiche per il caso originale dei biscotti ( $p = 1/8$ ) e per il caso in cui vi fosse un biscotto bruciacciato ogni 4 ( $p = 1/4$ ):

[binom6-1/8](#) e [binom6-1/4](#).



Con la nostra [grande CT](#) possiamo calcolare media e varianza delle due distribuzioni ([vedi](#)) ottenendo: nel caso di un biscotto bruciacciato ogni 8 media = 0.75, varianza = 0.65625, in quello di uno ogni 4 media = 1.5, varianza = 1.125. Dai grafici possiamo vedere la sensatezza dei due valori medi.

Posso osservare, come per le monete, che  $N$  è interpretabile come  $\sum_{i=1}^n X_i$  con  $X_i$  a valori in  $\{0, 1\}$  e distribuzione:  $\Pr(X_i=1) = p$ ,  $\Pr(X_i=0) = 1-p$ . Da qui, tenendo conto che

$$\bullet \mathbf{M}(X_i) = \Pr(X_i=0) \cdot 0 + \Pr(X_i=1) \cdot 1 = (1-p) \cdot 0 + p \cdot 1 = p \text{ e}$$

$$\bullet \mathbf{Var}(X_i) = \Pr(X_i=0) \cdot (0-p)^2 + \Pr(X_i=1) \cdot (1-p)^2 = (1-p)p^2 + p(1-p)^2 = p(1-p) \text{ abbiamo:}$$

$$M(N) = M(\sum_i X_i) = \sum_i M(X_i) = p + \dots + p = np \quad e:$$

$$Var(N) = Var(\sum_i X_i) = \sum_i Var(X_i) = p(1-p) + \dots + p(1-p) = np(1-p).$$

Nei due casi dei biscotti, in accordo coi valori calcolati sopra:

$$M(N) = 6 \cdot 1/8 = 3/4 = 0.75, \quad Var(N) = 6 \cdot 1/8 \cdot 7/8 = 21/32 = 0.65625$$

$$M(N) = 6 \cdot 1/4 = 3/2 = 1.5, \quad Var(N) = 6 \cdot 1/4 \cdot 3/4 = 9/8 = 1.125$$

Usualmente, la *legge binomiale* di ordine  $n$  e "probabilità di successo nella singola prova"  $p$  viene indicata con il simbolo  $B_{n,p}$ .

- 3** Un apparecchio, composto da 10 elementi omogenei, funziona a patto che funzionino almeno 7 elementi. L'*affidabilità* (cioè la probabilità di funzionamento perfetto) di ciascun elemento in un intervallo di tempo dato è 0.8. Qual è la probabilità che, nell'intervallo dato, l'apparecchio vada fuori uso?

**Nota.** La probabilità che ciascun elemento vada fuori uso è 0.2. Quindi devo considerare la legge binomiale  $B_{10,0.2}$  e valutare:

$\Pr(B_{10,0.2}=4) + \Pr(B_{10,0.2}=5) + \dots + \Pr(B_{10,0.2}=10)$ . Ma conviene calcolare:

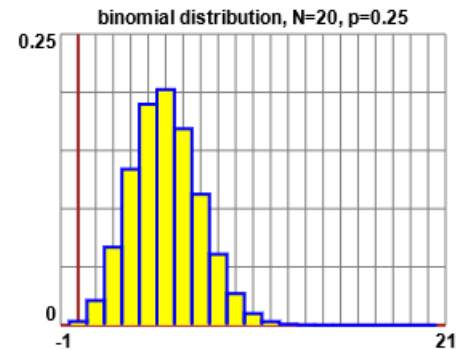
$1 - \Pr(\text{"apparecchio non va fuori uso"}) = 1 - (\Pr(B_{10,0.2}=0) + \dots + \Pr(B_{10,0.2}=3))$

## 2. La distribuzione gaussiana e il teorema limite centrale

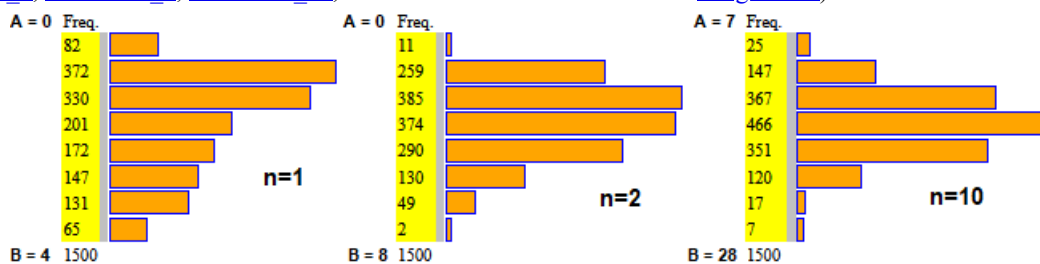
Sopra abbiamo visto la rappresentazione grafica della distribuzione binomiale per  $p = 0.25$  e  $n = 6$ . Qui a destra è rappresentato il caso in cui  $n$  è aumentato a  $n = 20$  ([vedi](#)).

Si vede che la forma dell'istogramma, al crescere di  $n$ , tende a stabilizzarsi sul grafico di una funzione da una forma particolare, simile a quella dell'esempio della durata delle telefonate studiato nella ➡ scheda *Quale matematica per i fenomeni casuali?*

Questo ci suggerisce che potremmo trovare un modo più semplice per calcolare le probabilità relative alle distribuzioni binomiali nei casi in cui  $n$  sia molto grande, anche per ovviare alle difficoltà di calcolo (anche col computer) che si generano.



Ricordiamo che ➡ la binomiale di ordine  $n$  è ottenibile come somma di  $n$  termini uguali ad una variabile casuale ad uscite in 0 ed 1. Vi sono altre variabili casuali che rappresentano la ripetizione di esperimenti e che all'aumentare del numero degli esperimenti tendono ad avere una legge di distribuzione il cui grafico tende ad assumere tale forma. Ad esempio consideriamo la somma di  $n$  termini pari a  $\sqrt{RND} + 3 \cdot RND^2$  (RND numero casuale con distribuzione uniforme tra 0 ed 1); lo studio per  $n=1$ ,  $n=2$  e  $n=10$ , con 1500 esperimenti (gli script: [varieRND\\_1](#), [varieRND\\_2](#), [varieRND\\_10](#), i cui esiti sono stati analizzati con [Istogramma](#)):



Prima di andare avanti, rivedete rapidamente il §5 della ➡ scheda *Quale matematica per i fenomeni casuali?*

Osservazioni e congetture fatte all'inizio del paragrafo possono essere precisate.

Siano  $U_i$  ( $i$  intero positivo) variabili casuali (numeriche) indipendenti con la stessa legge di distribuzione.

Al crescere di  $n$  la variabile casuale  $X_n = \sum_{i=1..n} U_i$  tende ad avere la legge di distribuzione, detta **normale** o **gaussiana**, con la seguente funzione di densità  $f$ , dove:

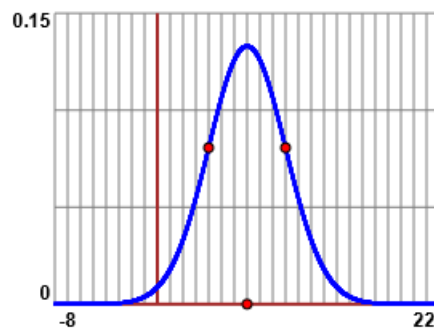
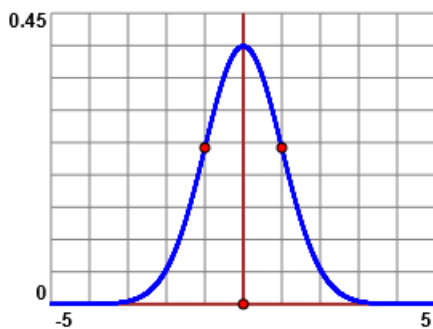
–  $m$  è la media, pari a  $n$  volte la media delle  $U_i$

–  $\sigma$  è lo scarto quadratico medio, pari a  $\sqrt{V}$  dove  $V$  è  $n$  volte la varianza delle  $U_i$  ( $\sigma$  è la lettera greca "sigma", che corrisponde alla nostra "s").

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\left(\frac{x-m}{\sigma}\right)^2 / 2} \quad \text{per } m=0 \text{ e } \sigma=1: \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

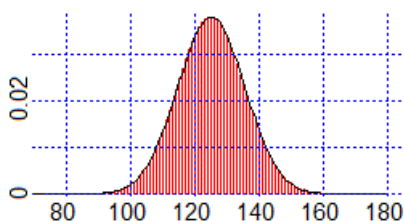
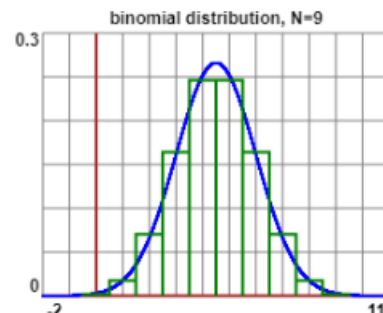
Questa proprietà è nota come **teorema limite centrale**.

Ecco il grafico di due "gaussiane", la prima con media 0 e  $\sigma=1$ , la seconda con media 7 e  $\sigma=3$ . Abbiamo evidenziato anche i punti dei grafici con input  $\text{media}-\sigma$  e  $\text{media}+\sigma$  (grafici realizzati con gli script [gauss](#) e [gauss1](#)).



Il nome "gaussiana" deriva da quello del matematico (o fisico, naturalista, filosofo, ...: le etichette attuali avevano un significato diverso un paio di secoli fa) **Gauss**, che la studiò particolarmente agli inizi dell'Ottocento; essa, in realtà, fu introdotta nel calcolo delle probabilità almeno una settantina d'anni prima.

Ecco, a destra, il grafico della gaussiana sovrapposto a quello dell'istogramma della distribuzione del ➡ quesito 2, avendo scelto la stessa media (4.5) e lo stesso sqm ( $\sqrt{2.25}$ ); il grafico è realizzato con lo script [BinomGauss](#).



Ecco, a sinistra, l'istogramma nel caso dei ➡ biscotti per una partita di 1000 pezzi e, sovrapposto, il grafico della gaussiana, con media  $N \cdot p = 125$  e scarto quadratico medio  $\sqrt{N \cdot p \cdot (1-p)} = 10.45825$ .

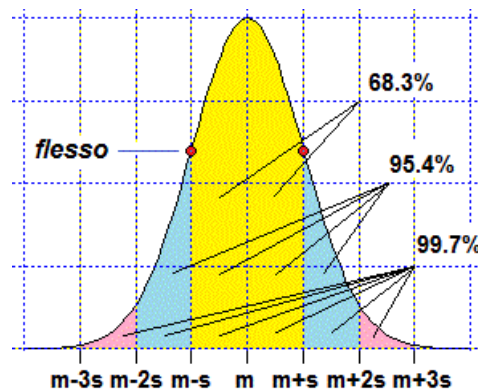
**4** Che cosa sono e quanto valgono, nell'esempio precedente,  $N$  e  $p$ ?

A lato è raffigurato il grafico di una generica *gaussiana* di media  $m$  e scarto quadratico medio  $s$ . Si può dimostrare che, oltre ad essere simmetrico rispetto alla retta di ascissa  $m$ , ha flessi (ossia cambi di concavità) nei punti di ascissa  $m-s$  e  $m+s$ .

Si può congetturare su vari esempi e dimostrare che l'integrale tra  $m-h \cdot s$  e  $m+h \cdot s$  di una densità gaussiana dipende solo da  $h$ , e non dai valori della media  $m$  e dello scarto quadratico medio  $s$ .

Ecco un esempio di calcolo (relativo all'immagine a destra) realizzato con lo script [IntGauss](#) che consente di integrare qualsiasi gaussiana:

```
0.99730020394 if a=162 b=198, m=180 sigma=6
0.95449973539 if a=168 b=192, m=180 sigma=6
0.68268949234 if a=174 b=186, m=180 sigma=6
```



Non c'è un termine semplice (esprimibile mediante quattro operazioni, funzioni polinomiali, circolari, esponenziali o loro inverse) con cui rappresentare l'integrale di una gaussiana.

**5** Quanto vale l'area tra curva gaussiana ed asse  $x$  a destra della retta  $y = m$  (se  $m$  è il valor medio)?

### 3. Dalla media statistica alla media teorica

Voglio determinare il peso medio della popolazione adulta (di un certo paese) di un dato sesso, ad es. maschile. Ovvero, se  $P$  è la variabile casuale "peso di un abitante adulto maschio", voglio determinare  $M(P)$ . Indico con  $\sigma$  lo sqm di  $P$ . Rilevo i pesi  $P_1, P_2, \dots, P_n$  di un *campione* di  $n$  persone.

$\sum_i P_i / n$  ( $i=1..n$ ) viene chiamata *media statistica* di  $P$  di ordine  $n$ ; indichiamola con  $M_n(P)$ . Anch'essa è una variabile casuale: a seconda degli  $n$  soggetti che considero ottengo valori leggermente diversi. Le  $P_i$  sono tutte variabili casuali distribuite come  $P$  (se prendo le persone in modo del tutto casuale); se faccio i rilevamenti in modo indipendente, per il teorema limite centrale ho che  $\sum_i P_i$  al crescere di  $n$  tende ad avere andamento gaussiano con media  $n M(P)$  e varianza  $n \text{Var}(P)$ , ovvero scarto quadratico medio  $\sqrt{n} \sigma$ .

Dividendo per  $n$  ho  $M_n(P) = \sum_i P_i / n$  che, quindi, al crescere di  $n$ , **tende ad avere andamento gaussiano con media  $M(P)$  e sqm  $\sigma/\sqrt{n}$** . Lo sqm di questa gaussiana tende a 0, per cui il valore  $M_n(P)$  che ottengo tende a cadere sempre più vicino a  $M(P)$ .

Ad es., supponendo che  $P$  sia espressa in kg, se voglio determinare il peso medio della popolazione a meno di 0.5 kg posso fare tante prove  $n$  fino a che  $\sigma/\sqrt{n} < 0.5$ . A quel punto potrò dire che, con ➡ probabilità del 68.3%, il valore  $M_n(P)$  trovato approssima  $M(P)$  a meno di 0.5. Se voglio una stima più sicura, praticamente certa, posso fare tante prove  $n$  fino a che  $3\sigma/\sqrt{n} < 0.5$ . A quel punto potrò dire che, con probabilità del 99.7%, il valore  $M_n(P)$  trovato dista da  $M(P)$  meno di 0.5.

Quanto qui detto per  $P$  vale per ogni variabile casuale.

Il valore di  $\sigma$  devo già conoscerlo in base a considerazioni di qualche tipo oppure posso man mano approssimarlo con la radice quadrata della varianza sperimentale: si può dimostrare che, fissato  $n$ , la varianza di  $M_n(X)$ , calcolata ripetutamente, dà luogo a valori la cui media tende a  $\text{Var}(X) \cdot (n-1)/n$ .

Ovvero come  $\sigma$  devo prendere il *valore sperimentale* ( $\Rightarrow$ ) *moltiplicato per*  $\sqrt{(n/(n-1))}$ . Ovvero devo prendere il secondo dei valori che, in modo non molto corretto ma ormai diffuso, vengono in genere indicati nel modo seguente (dove  $x_i$  e  $\mu$  sono dati e media):

$$\sigma_n = \left( \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2}{n} \right)^{1/2}$$

$$\sigma_{n-1} = \left( \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2}{n-1} \right)^{1/2}$$

Questi due termini vengono, spesso, chiamati, rispettivamente, *deviazione standard teorica* e *deviazione standard corretta o non distorta* o *statistica* o *sperimentale*. Spesso vengono entrambi chiamati semplicemente **deviazione standard**. Sta al lettore capire quale uso si sta facendo. Comunque quando  $n$  è abbastanza grande i due numeri hanno una piccola differenza relativa. Nella nostra **grande CT** sono presenti tre tasti, di cui ora chiariamo il significato:

n max min sum mul median mean var sqr sd sigma 3sigma C

10, 11, 12, 13, 14, 15, 16, 17, 18, 19

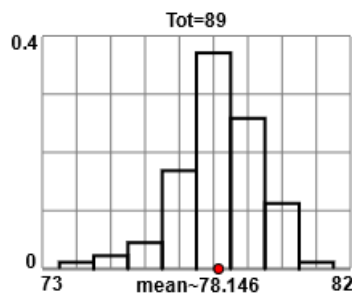
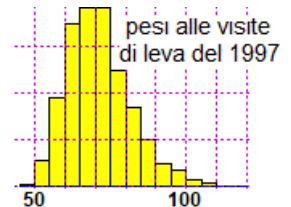
scarto quad. medio (sq.root of var./theoret.st.dev.) = 2.8722813232690143  
experimental standard dev. = 3.0276503540974917  
sigma = 0.9574271077563381

**[sqr]** calcola lo scarto quadratico medio o deviazione standard teorica, **[sd]** calcola la deviazione standard sperimentale, **[sigma]** calcola la **[sd]** diviso per  $\sqrt{n}$ , ossia lo **sqr** della media dei dati.

**6** Prova a calcolare **[sqr]** e **[sd]** dei 100, dei 1000 e dei 10000 dati seguenti. Che cosa osservi?

10\*10, 11\*10, 12\*10, 13\*10, 14\*10, 15\*10, 16\*10, 17\*10, 18\*10, 19\*10  
10\*100, 11\*100, 12\*100, 13\*100, 14\*100, 15\*100, 16\*100, 17\*100, 18\*100, 19\*100  
10\*1e3, 11\*1e3, 12\*1e3, 13\*1e3, 14\*1e3, 15\*1e3, 16\*1e3, 17\*1e3, 18\*1e3, 19\*1e3

Ricordiamo (facendo riferimento all'esempio iniziale del paragrafo) che è la *media* dei pesi che si misurano ad avere *andamento gaussiano*, non i pesi stessi. Per confermare questo si consideri l'istogramma a lato, che rappresenta la distribuzione dei pesi (in kg) rilevati alle visite di leva per la Marina del 1997 (primi scaglioni); si tratta di circa 4 mila maschi italiani ventenni.



Questi strumenti entrano in gioco anche nel caso di alcuni tipi di misurazioni. Da una parte abbiamo quelle a **bassa sensibilità**, che, se ripetute, danno luogo agli stessi valori. Un esempio: se uso un doppio decimetro per misurare la lunghezza  $L$  di un oggetto, individuo la tacca più vicina alla estremità dell'oggetto; se ad es. questa rappresenta 13.4 cm, prendo  $13.4 \text{ cm} \pm 1/2 \text{ mm}$  come approssimazione *certa* di  $L$ , cioè  $[13.35, 13.45]$  come intervallo di indeterminazione per il valore in cm di  $L$ .

Dall'altra vi sono quelle ad **alta sensibilità**. Un esempio: consideriamo un particolare riflettometro per la determinazione del tasso glicemico del sangue (viene depositata una goccia di sangue su una striscetta, che viene inserita nello strumento; questo, mediante un opportuno dispositivo ottico, effettua il rilevamento ed esprime il valore in mg/dl); in dotazione è presente (per verificare il corretto funzionamento dell'apparecchio) una striscia di controllo trattata in modo da avere caratteristiche corrispondenti a quelle di una goccia di sangue con un certo tasso glicemico. Se ripeto più volte la misurazione del "tasso glicemico" della striscia di controllo, anche se esso non cambia, non ottengo sempre lo stesso valore. A sinistra è riprodotto l'istogramma relativo a 89 rilevamenti.

Le variazioni tra un rilevamento e l'altro sono dovute a una serie di fattori casuali (in cui intervengono aspetti ottici, elettronici, ) che non è possibile né eliminare né conoscere esattamente, per cui il valore letto è da interpretare come una variabile casuale. Aumentando il numero delle prove l'istogramma tende a stabilizzarsi.

Qual è il tasso glicemico della striscia di controllo? Con la **grande CT**, introducendo i valori rilevati (quelli sotto riportati) ottengo:

74, 75\*2, 76\*4, 77\*15, 78\*33, 79\*23, 80\*10, 81  
mean = 78.14606741573034  
sigma = 0.13133882936711402  
3\*sigma = 0.39401648810134204

Con probabilità del 99.7% il valore atteso della variabile "esito di un rilevamento per la striscia di controllo" è  $78.146 \pm 0.394$ . Questo non è un intervallo di indeterminazione "certo".

Ovviamente questo era un esempio. Se faccio un rilevamento per una certa persona non ha senso effettuare molte prove e calcolare il valor medio: in pochissimo tempo il tasso può cambiare, per cui ripetendo la prova non è detto che si effettui sempre la misurazione della stessa grandezza. E, poi, non interessano valutazioni molto precise. Tenendo conto dello studio effettuato sulla striscia di controllo (con uscite tra 74 e 81) possiamo stimare che anche in altre situazioni lo scarto che il valore letto può avere dal valore vero sia inferiore a 4 e associare alla misura letta tale precisione: se leggo 93, assumo che il tasso glicemico sia  $93 \pm 4 \text{ mg/dl}$ .

Nel caso di misurazioni di altri fenomeni in genere il valore individuato non cambia ogni volta che si effettua un rilevamento. Un esempio:

**7** Con un apparato misuratore ad alta sensibilità si ottengono le 7 misure (in un'opportuna unità di misura): 7.3, 7.1, 7.2, 6.9, 7.2, 7.3, 7.4. Determina un intervallo in cui al 68.3% cada il "valore vero" della misura. Determinane uno in cui cada al 99.7%.

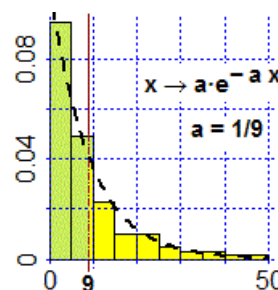
#### 4. La distribuzione esponenziale



La durata delle telefonate ➡ per la ditta Telstat considerata nella scheda *Quale matematica per i fenomeni casuali?* aveva approssimativamente andamento gaussiano. Vediamo, ora, con quale funzione possiamo approssimare la distribuzione dei tempi tra una telefonata e l'altra. A destra è riprodotto un tale istogramma di distribuzione (le ascisse rappresentano gli intervalli di tempo in secondi) e, sovrapposto, tratteggiato, il grafico di una funzione che sembra approssimarla. Il tempo medio di attesa era di 9 secondi.

Fenomeni di questo tipo (come ad esempio anche la distanza temporale tra la venuta al semaforo di un'auto e la venuta dell'auto successiva, nel caso di un semaforo preceduto da un lungo tratto di strada senza impedimenti) hanno una *distribuzione*, chiamata *esponenziale*, che ha come funzione di densità la seguente, dove  $a$  è il reciproco della media (nel nostro caso  $a = 1/9$ ):

$$x \rightarrow a \cdot e^{-ax} \quad (x > 0)$$



**8** Per una particolare valore di  $M$ ,  $M = 9$ , posto  $A = 1/M$ , avendo definito  $F(x) = A \cdot \exp(-A \cdot x)$ , si sono calcolati in modo approssimato (usando gli script [integ\\_exp1](#), [integ\\_exp2](#), [integ\\_exp3](#)):

$$\int_0^\infty F(x) dx \quad [=100\%] \quad \int_0^\infty x \cdot F(x) dx \quad [=mean] \quad \int_0^\infty (x-M)^2 \cdot F(x) dx \quad [=sqm^2]$$

Che cosa puoi dedurre sui valori della media e dello scarto quadratico medio della distribuzione esponenziale?

```
F(x) = A*exp(-A*x)
mean=9; A=1/mean; y=A*Math.exp(-A*x)
0.9999999994852756 if a=0 b=1000 n=1e6          0 ∫ F(x) dx = 1
mean=9; A=1/mean; y=A*Math.exp(-A*x); y=x*y
9.000000004624004 if a=0 b=1000 n=1e6          0 ∫ x * F(x) dx = 9
mean=9; A=1/mean; y=A*Math.exp(-A*x); y=(x-mean)*(x-mean)*y
80.9999998749507 if a=0 b=1000 n=1e6          0 ∫ (x-mean)^2 * F(x) dx = 81
```

## 5. Approfondimenti

### La legge dei grandi numeri

Per inciso, osserviamo che alcune considerazioni svolte nei paragrafi precedenti sono descritte mediante proprietà note come *leggi dei grandi numeri*, e a volte raggruppate sotto la voce *legge di Bernoulli* in quanto Jakob Bernoulli - intorno al 1700 - ne dette una prima formulazione. Possiamo dare una descrizione sintetica di questa legge dicendo che se  $U_1, \dots, U_n$  sono  $n$  variabili casuali con la stessa legge di distribuzione, di media  $\mu$ , allora  $\sum_{i=1}^n U_i / n$  *converge in probabilità* a  $\mu$ , ossia, fissata comunque una probabilità  $P$ , per ogni  $\varepsilon > 0$  posso trovare  $n$  tale che, da lì in poi,  $\sum_{i=1}^n U_i / n$  disti da  $\mu$  meno di  $\varepsilon$  con probabilità  $P$ .

Il concetto di "*limite in probabilità*", ora descritto, coincide con quello usuale, a parte il fatto che si trova un valore di  $n$  a partire dal quale vale la disuguaglianza non con certezza, ma con una certa probabilità. Ciò corrisponde al fatto che, ad es., se lancio una coppia di dadi prima o poi la media delle uscite si stabilizza attorno a 7, ma, anche se è altamente improbabile, potrebbe accadere che a un certo punto si susseguano 20 uscite uguali a 2 che abbassino, provvisoriamente, la media.

### Sulla legge di distribuzione binomiale simmetrica

Abbiamo visto ➡ che la distribuzione binomiale simmetrica ( $p=1/2$ ) ad  $n$  prove ha *media*  $n/2$ . Abbiamo congetturato che la *varianza* sia  $n/4$  e che, quindi, lo scarto quadratico medio sia  $\sigma = \sqrt{n}/2$ . La cosa può essere effettivamente dimostrata. Sia  $N$  la variabile casuale:

- $N = u_1 + u_2 + \dots + u_n$  dove per tutti gli  $u_k$  si ha  $u_k=1$  se al primo lancio è uscita T,  $u_k=0$  altrimenti.
- i lanci sono indipendenti, quindi ➡ la "varianza della somma è la somma delle varianze";
- $M(u_k) = 1/2$ ;  $\text{Var}(u_k) = 1/2 (0-1/2)^2 + 1/2 (1-1/2)^2 = 1/4$ ; ...
- $\text{Var}(N) = \text{Var}(u_1) + \dots + \text{Var}(u_n) = 1/4 \cdot n = n/4$

## 6. Esercizi

- e1** Lanciamo una moneta equa 10 volte. Qual è la probabilità che testa venga più frequentemente di croce?
- e2** In un gran numero di esperimenti si ottiene che un certo veleno uccide l'80% dei topi su cui viene provato. Se viene usato su un gruppo di 5 topi, quali sono le probabilità che 0, 1, 2, 3, 4, 5 topi siano uccisi? [arrotonda i valori ai centesimi di percentuale]
- e3** Voglio mangiare uno confetto col cuore di mandorla. Ho una confezione formata da 3 confetti col cuore di cioccolata e 3 col cuore di mandorla. Quanti confetti "mi attendo" di dover mangiare per trovarne uno col cuore di mandorla?
- e4** Una partita di semi ha un tasso di germinazione dell'80%. Se si piantano due semi in una buca, indicando con  $N$  il numero di piante che germinano, quanto valgono (a)  $\text{Pr}(N=0)$ , (b)  $\text{Pr}(N=1)$ , (c)  $\text{Pr}(N=2)$  e (d) il valore atteso  $M(N)$ ?
- e5** Una azienda produce chip dei quali il 10% sono difettosi. Se vengono acquistati 100 chip da tale azienda, il numero dei chip difettosi che si trovano è una variabile casuale binomiale o lo è solo sotto opportune ipotesi?
- e6** Calcola  $\text{Pr}(-1.2 \leq U \leq 1.6)$  nel caso in cui  $U$  abbia distribuzione gaussiana con  $M(U) = 1.2$  e  $\sigma(U) = 0.8$ .
- e7** Da una estesa analisi risulta che i maschi quarantenni in un certo paese hanno altezze normalmente distribuite con media 175 cm e sc. quad. medio 5 cm. In un campione casuale di 100 maschi quarantenni di quel paese quanti te ne aspetti di altezza tra i 170 e i 180 centimetri? Quanti di altezza superiore a 165 centimetri?
- e8** Sia  $X$  una variabile casuale continua a valori in  $[0, \infty)$  con funzione di densità  $x \rightarrow 7e^{-hx}$ . Quanto vale  $h$ ? Quanto valgono media e scarto quadratico medio?

1) Segna con l'evidenziatore, nelle parti della scheda indicate, frasi e/o formule che descrivono il significato dei seguenti termini:

*legge di distribuzione di Bernoulli (§1), legge di distribuzione normale (§2), teorema limite centrale (§2), deviazione standard (§3), legge di distribuzione esponenziale (§4).*

2) Su un foglio da "quadernone", nella prima facciata, esemplifica l'uso di ciascuno dei concetti sopra elencati mediante una frase in cui esso venga impiegato.

3) Nella seconda facciata riassumi in modo discorsivo (senza formule, come in una descrizione "al telefono") il contenuto della scheda (non fare un elenco di argomenti, ma cerca di far capire il "filo del discorso").

**script:** [piccola CT](#) [grande CT](#) [isto](#) [isto con %](#) [boxplot](#) [striscia](#) [100](#) [ordina](#) [Grafici](#) [GraficD](#) [divisori](#) [Indet](#) [distanza](#) [Triang](#)  
[eq.polinomiale](#) [eq.nonPolin](#) [sistemaLin](#) [moltPolin](#) [sempliciEq](#) [divisori](#) [fraz/mcd](#) [opFraz](#) [SumPro](#) [sin](#) [LenArc](#) [Poligono](#)  
[Circ3P](#) [Inscr3P](#) [IntegrPol](#) [Istogramma](#) [RandomNum](#) [binom10](#) [binom9](#) [binom6-1/8](#) [binom6-1/4](#) [binom20](#) [varieRND\\_1](#)  
[varieRND\\_2](#) [varieRND\\_10](#) [gauss](#) [gauss1](#) [BinomGauss](#) [IntGauss](#) [integ\\_exp1](#) [integ\\_exp2](#) [integ\\_exp3](#)