www.sce.unimore.it

Scienze della Comunicazione
e dell'Economia

NUMERI ALEATORI E LORO DISTRIBUZIONI

Luca La Rocca
luca.larocca@unimore.it

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Un numero aleatorio (variabile casuale) è una quantità il cui valore dipende dell'esito di un esperimento aleatorio.

L'esempio più semplice è l'indicatrice di un evento:

$$\mathbb{I}_E = \begin{cases} 1 & \text{se } E \text{ è VERO,} \\ 0 & \text{se } E \text{ è FALSO.} \end{cases}$$

Altri esempio sono:

- il punteggio ottenuto con il lancio di uno o più dadi;
- l'altezza di uno studente preso a caso dalla sua classe;
- il reddito medio di un campione casuale di famiglie italiane;
- la durata della vita di un individuo.

Formalmente, se Ω è l'insieme ambiente dell'esperimento aleatorio da cui dipende il numero aleatorio X, questo è definito come una funzione che a ogni esito $\omega \in \Omega$ associa il numero reale $X(\omega) \in \mathbb{R}$.

Per esempio, con riferimento al lancio di due dadi a quattro facce, se

$$\begin{aligned}\Omega = \{&(1,1),(1,2),(1,3),(1,4),\\ &(2,1),(2,2),(2,3),(2,4),\\ &(3,1),(3,2),(3,3),(3,4),\\ &(4,1),(4,2),(4,3),(4,4)\}\end{aligned}$$

è l'insieme ambiente...

…sono numeri aleatori

- il punteggio ottenuto con il primo dado,

$$X_1(\omega) = \omega_1,$$

- il punteggio ottenuto con il secondo dado,

$$X_2(\omega) = \omega_2,$$

- il punteggio ottenuto con (la somma de)i due dadi,

$$S(\omega) = X_1(\omega) + X_2(\omega) = \omega_1 + \omega_2,$$

dove $\omega = (\omega_1, \omega_2)$ è la generica coppia di punteggi; es. $X_1(2,3) = 2$, $X_2(2,3) = 3$ e $S(2,3) = 5$.

Analogamente, se $\omega = (\omega_1, \ldots, \omega_n)$ denota un campione casuale di famiglie italiane, il loro reddito medio sarà definito come

$$\overline{R}(\omega) = \frac{R_1(\omega) + \cdots + R_n(\omega)}{n}$$

dove R_i è il reddito dichiarato dall'i-esima famiglia intervistata, $i = 1, \ldots, n$.

Scriveremo compattamente $S = X_1 + X_2$ e

$$\overline{R} = \frac{R_1 + \cdots + R_n}{n}$$

quando riterremo di potere sottintendere la dipendenza da $\omega \in \Omega$.

Si dice distribuzione di un numero aleatorio la probabilità che questo individua sui plurintervalli della retta reale.

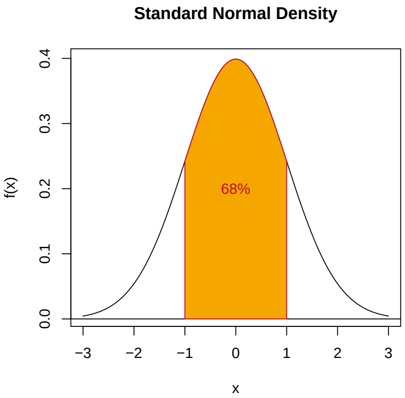
Per esempio, se X_1 è il punteggio ottenuto con un dado a quattro facce, supposto bilanciato, si trova

$$
\begin{array}{rcrcl}
\mathbb{P}\{1 < X_1 \leq 4\} & = & \mathbb{P}\{2,3,4\} & = & 0.75 \\
\mathbb{P}\{X_1 \geq 3\} & = & \mathbb{P}\{3,4\} & = & 0.50 \\
\mathbb{P}\{X_1 < 1.5\} & = & \mathbb{P}\{1\} & = & 0.25 \\
\mathbb{P}\{X_1 = 1.5\} & = & \mathbb{P}\{\} & = & 0.00 \\
\mathbb{P}\{X_1 \neq 1.5\} & = & \mathbb{P}\{1,2,3,4\} & = & 1.00 \\
\mathbb{P}\{X_1 \leq 2.5 \text{ o } X_1 \geq 4.5\} & = & \mathbb{P}\{1,2\} & = & 0.50 \\
\mathbb{P}\{X_1 < 2 \text{ o } X_1 = 3 \text{ o } X_1 \geq 4\} & = & \mathbb{P}\{1,3,4\} & = & 0.75
\end{array}
$$

Se invece A è l'altezza in unità standard di uno studente estratto a caso da una classe di cui siano note la media e la deviazione standard, si trova per esempio

$$\begin{aligned} \mathbb{P}\{-1 \le A \le 1\} &= 0.68 \\ \mathbb{P}\{A \ge 1\} &= 0.16 \\ \mathbb{P}\{A \le -1\} &= 0.16 \\ \mathbb{P}\{A \ge -1\} &= 0.84 \end{aligned}$$

calcolando numericamente le corrispondenti aree sotto la funzione di densità normale standard (supposta adeguata).



$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \\
\mathbb{P}(-1,1) &= \int_{-1}^{1} f(x)\,dx \simeq 0.68 \\
\mathbb{P}(1,\infty) &= \int_{1}^{\infty} f(x)\,dx \simeq 0.16 \\
\mathbb{P}(-\infty,-1) &= \int_{-\infty}^{-1} f(x)\,dx \simeq 0.16 \\
\mathbb{P}(-1,\infty) &= \int_{-1}^{\infty} f(x)\,dx \simeq 0.84
\end{aligned}
$$

Si dice funzione di probabilità di un numero aleatorio

l'elenco dei singoli valori che questo può assumere con probabilità positiva e delle probabilità con cui tali valori, detti atomi, vengono assunti.

Se l'unione degli atomi ha probabilità uno (come nel caso della somma dei punteggi di due dadi a quattro facce) si parla di numero aleatorio discreto e la funzione di probabilità determina l'intera distribuzione:

la probabilità di un plurintervallo sarà la somma delle probabilità degli atomi in esso contenuti.

Per esempio, se S è la somma dei punteggi di due dadi a quattro facce, supposti <u>bilanciati</u>, si trova

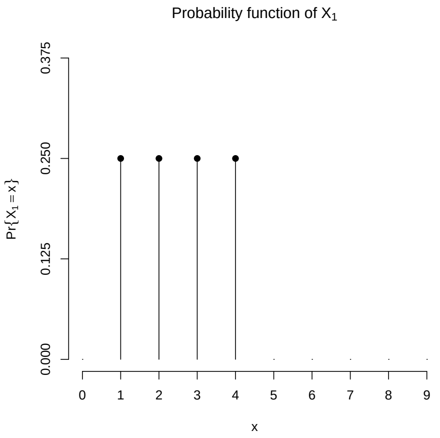
$$
\begin{aligned}
\mathbb{P}\{S=2\} &= \mathbb{P}\{(1,1)\} &= 1/16\\
\mathbb{P}\{S=3\} &= \mathbb{P}\{(1,2),(2,1)\} &= 2/16\\
\mathbb{P}\{S=4\} &= \mathbb{P}\{(1,3),(2,2),(3,1)\} &= 3/16\\
\mathbb{P}\{S=5\} &= \mathbb{P}\{(1,4),(2,3),(3,2),(4,1)\} &= 4/16\\
\mathbb{P}\{S=6\} &= \mathbb{P}\{(2,4),(3,3),(4,2)\} &= 3/16\\
\mathbb{P}\{S=7\} &= \mathbb{P}\{(3,4),(4,3)\} &= 2/16\\
\mathbb{P}\{S=8\} &= \mathbb{P}\{(4,4)\} &= 1/16
\end{aligned}
$$

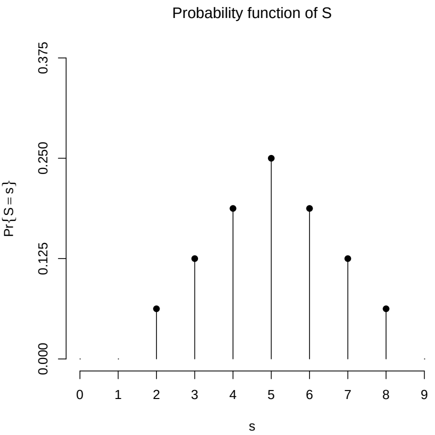
e quindi $\mathbb{P}\{4 < S \le 6\} = \mathbb{P}\{S=5\} + \mathbb{P}\{S=6\} = 7/16$ (etc).

Una funzione di probabilità può essere rappresentata graficamente mediante un grafico a bastoncini: si individuano gli atomi su un asse orizzontale e, in corrispondenza di ciascun atomo, si traccia un segmento verticale di altezza pari alla corrispondente probabilità.

Spesso, per conseguire una maggiore efficacia grafica, si aggiunge un “pallino” in testa ai segmenti (a rappresentare concretamente la “massa” di probabilità portata dall’atomo).

Le figure seguenti rappresentano le funzioni di probabilità di X_1 ed S...


Probability function of X_1
0.375
0.250
0.125
0.000
$\Pr\{X_1 = x\}$
0
1
2
3
4
5
6
7
8
9
x


Probability function of S
Pr{S = s}
0.000
0.125
0.250
0.375
0
1
2
3
4
5
6
7
8
9
s

Se la distribuzione di un numero aleatorio è assegnata mediante una funzione di densità sulla retta reale, $f : \mathbb{R} \to \mathbb{R}_+$ con $\int_{-\infty}^{\infty} f(x)dx = 1$, come quella normale standard (nell'esempio dell'altezza in unità standard di uno studente estratto a caso da una classe di cui siano note la media e la deviazione standard) allora non ci sono atomi, perché tutti i singoli valori che A può assumere hanno probabilità nulla (i valori della funzione di densità sono probabilità per unità di lunghezza):

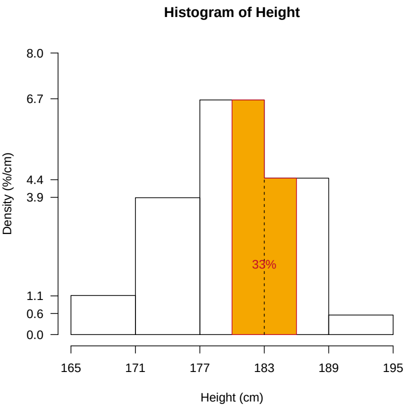
$$\mathbb{P}\{A = a\} \quad = \quad \int_a^a f(x)\,dx \quad = \quad 0, \qquad \text{per ogni } a \in \mathbb{R};$$

si parla in questo caso di numero aleatorio continuo.

Per numeri aleatori continui la funzione di probabilità è inutile.

Se, invece di essere note la media e la deviazione standard di classe, è nota la distribuzione di frequenza relativa a una suddivisione in classi, possiamo utilizzare il corrispondente istogramma come funzione di densità:

$$\begin{aligned}\mathbb{P}\{180 \leq A \leq 186\} &= 0.067 \times 3 + 0.044 \times 3 \\ &= 0.201 + 0.132 \\ &= 0.333\end{aligned}$$



Se a una classe di 30 studenti si aggiunge un insegnante alto 192 cm, fermo restando che le nostre conoscenze sulle altezze degli studenti sono descritte dall'istogramma precedente, avremo un unico atomo

$$\mathbb{P}\{A = 192\} \quad = \quad \frac{1}{31} \quad \simeq \quad 3.2\%$$

e l'istogramma andrà riscalato in modo da avere area pari a

$$\mathbb{P}(\text{“studente”}) \quad = \quad \frac{30}{31} \quad \simeq \quad 96.8\%.$$

Troveremo dunque...

...come la regola delle probabilità totali conferma:

$$
\begin{aligned}
\mathbb{P}\{180 \leq A \leq 186\} &= \frac{30}{31} \times 0.333 + \frac{1}{31} \times 0 \simeq 32\%, \\
\mathbb{P}\{189 \leq A \leq 195\} &= \frac{30}{31} \times 0.006 \times 6 + \frac{1}{31} \times 1 \simeq 6.7\%.
\end{aligned}
$$

Si parla in questo caso di numero aleatorio misto. Un altro esempio può essere il tempo di attesa all'ufficio postale: nullo se al nostro arrivo la coda è vuota, evento che si verifica con probabilità non nulla, anche se magari piccola, altrimenti pari al tempo necessario per servire i clienti davanti a noi (dunque privo di atomi).

Infine, se è nota la distribuzione unitaria dell'altezza nella classe (o comunque la distribuzione di frequenza relativa alle singole modalità osservate) avremo un altro esempio di numero aleatorio discreto:

- i suoi atomi saranno dati dalle modalità osservate;
- i corrispondenti valori della funzione di probabilità sarano dati dalle rispettive frequenze relative.

Si noti come la distribuzione di un numero aleatorio (es. altezza di uno studente preso a caso dalla sua classe) e quindi il suo essere continuo, discreto o misto, dipenda in modo essenziale dalle informazioni in possesso del soggetto che effettua la valutazione di probabilità.

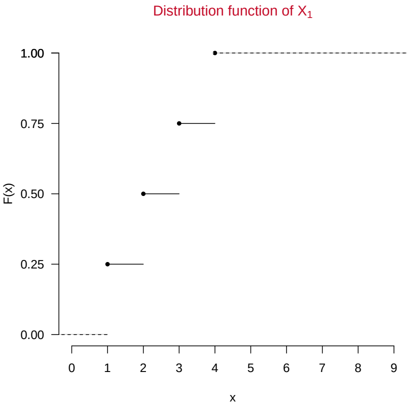
La funzione di ripartizione del numero aleatorio X è definita come

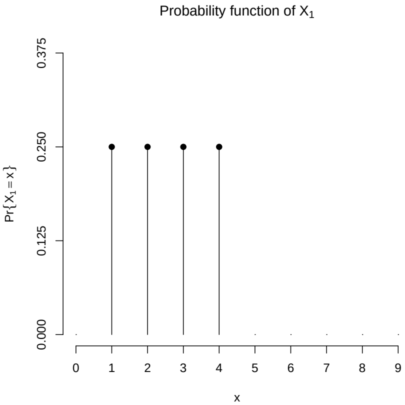
$$F(x) = \mathbb{P}\{X \leq x\}, \quad x \in \mathbb{R},$$

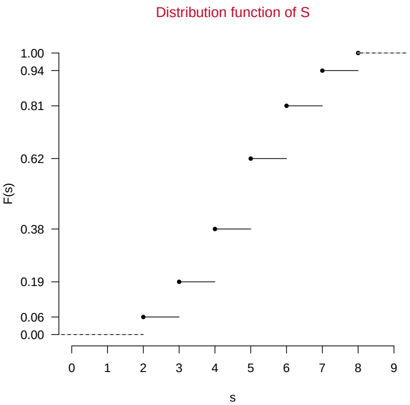
ovvero come la probabilità che X non superi x, al variare di $x \in \mathbb{R}$; ne consegue che $\mathbb{P}\{y < X \leq x\} = F(x) - F(y)$ per ogni x e y.

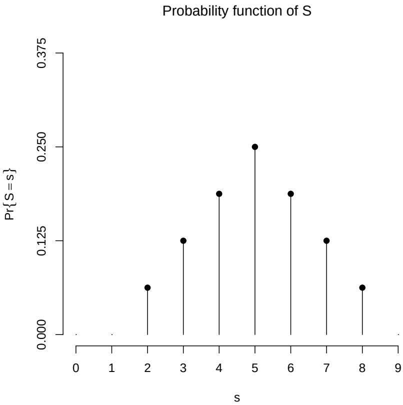
Per esempio, con riferimento al lancio di due dadi bilanciati a quattro facce, le figure seguenti riportano i grafici delle funzioni di ripartizione di X_1 (punteggio ottenuto con il primo dado) ed $S = X_1 + X_2$ (punteggio ottenuto con entrambi i dadi)...

Funzione di ripartizione





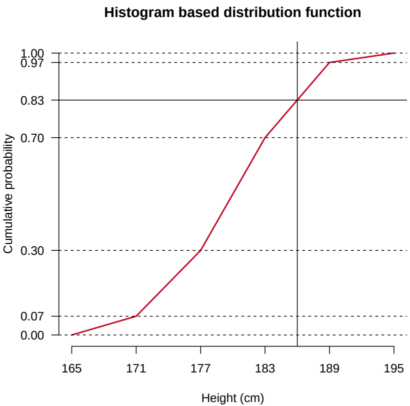

Distribution function of S
1.00
0.94
0.81
0.62
0.38
0.19
0.06
0.00
F(s)
0 1 2 3 4 5 6 7 8 9
s


Probability function of S
0.375
0.250
0.125
0.000
Pr{S = s}
0 1 2 3 4 5 6 7 8 9
s

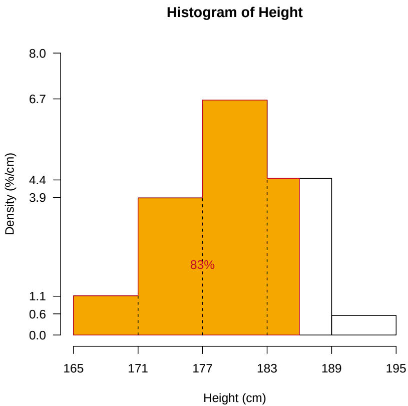
…si noti che le funzioni di ripartizione rappresentate nelle precedenti figure crescono solo “saltando” laddove il numero aleatorio presenta un atomo: questo è caratteristico dei numeri aleatori discreti.

Le ampiezze dei salti sono pari alle probabilità degli atomi e, assieme alle loro posizioni, individuano la funzione di probabilità.

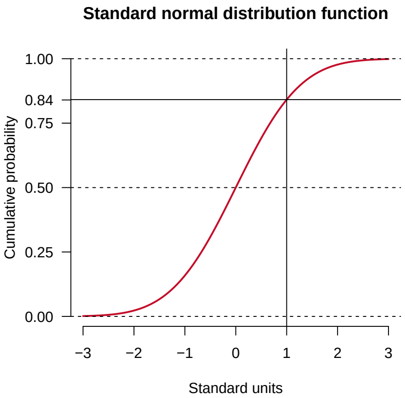
Le figure seguenti mostrano invece due funzioni di ripartizione (la normale standard e quella associata all’istogramma delle altezze precedentemente esibito) che non “saltano” mai; questo è caratteristico dei numeri aleatori continui…

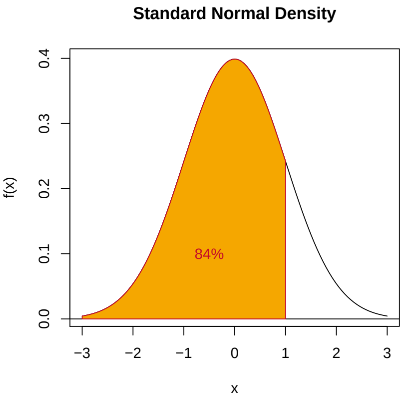


Histogram based distribution function
Cumulative probability
1.00
0.97
0.83
0.70
0.30
0.07
0.00
165
171
177
183
189
195
Height (cm)



Histogram of Height
Density (%/cm)
8.0
6.7
4.4
3.9
1.1
0.6
0.0
83%
165
171
177
183
189
195
Height (cm)


Standard normal distribution function
1.00
0.84
0.75
0.50
0.25
0.00
Cumulative probability
−3 −2 −1 0 1 2 3
Standard units


Standard Normal Density
0.4
0.3
0.2
0.1
0.0
f(x)
84%
−3 −2 −1 0 1 2 3
x

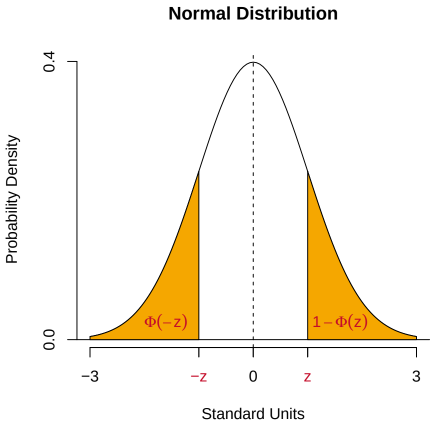
I valori della funzione di ripartizione normale standard, Φ, possono ottenersi numericamente (avvalendosi di un calcolatore) o mediante una tabella (avvalendosi di un libro).

Borra & Di Ciaccio (2004, p. 455) tabulano $\Phi(z)$ per $0 \leq z \leq 4$, con due cifre decimali per z e quattro cifre decimali per $\Phi(z)$: es. $\Phi(1.00) = 0.8413$.

Se si vuole trovare $\Phi(-z)$, con $-z < 0$, si sfrutta la simmetria

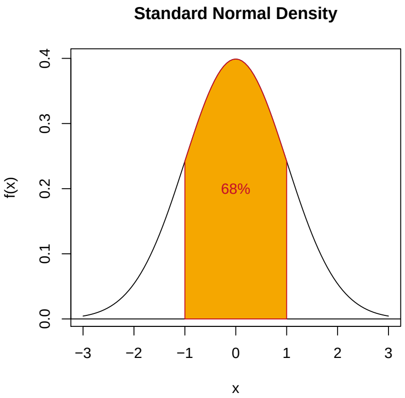
$$\Phi(-z) = 1 - \Phi(z)$$

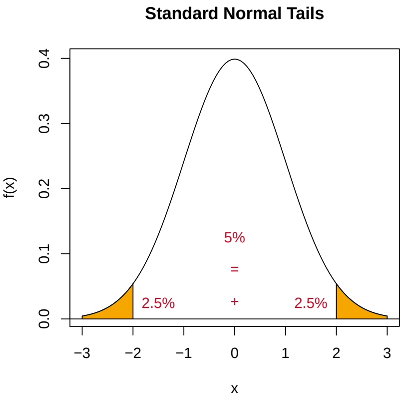
es. $\Phi(-1.00) = 1 - \Phi(1.00) = 1 - 0.8413 = 0.1587\ldots$


Normal Distribution
Probability Density
0.4
0.0
Φ(−z)
1 − Φ(z)
−3
−z
0
z
3
Standard Units

...in questo modo la tabella di Borra & Di Ciaccio (2004, p. 455) consente di calcolare la probabilità di un qualsiasi plurintervallo (espresso in unità standard); per esempio troviamo

$$\begin{aligned}
\mathbb{P}\{-1 \leq Z \leq 1\} &= \Phi(1) - \Phi(-1) \\
&= 0.8413 - 0.1587 \\
&= 0.6826 \\
\mathbb{P}\{Z < -2 \text{ } o \text{ } Z > 2\} &= \Phi(-2) + (1 - \Phi(2)) \\
&= (1 - 0.9772) + (1 - 0.9772) \\
&= 2 \times 0.0228 \\
&= 0.0456
\end{aligned}$$


Standard Normal Density
f(x)
0.0
0.1
0.2
0.3
0.4
68%
−3
−2
−1
0
1
2
3
x

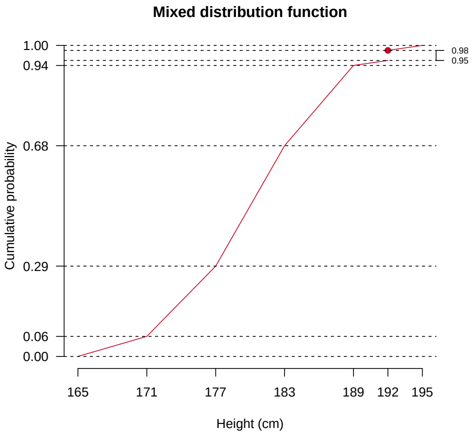

Standard Normal Tails
f(x)
0.0
0.1
0.2
0.3
0.4
5%
=
2.5%
+
2.5%
−3
−2
−1
0
1
2
3
x

In generale la funzione di ripartizione di un numero aleatorio misto crescerà sia “saltando”, in corrispondenza dei suoi atomi, sia “senza saltare” (tra un atomo e l’altro).

Anche in questo caso sarà possibile ricostruire la funzione di probabilità a partire dai salti, ma questa non descriverà completamente la distribuzione del numero aleatorio (ne descriverà soltanto gli atomi).

Si noti la continuità da destra del grafico.

Funzione di ripartizione



Il valore atteso (speranza matematica, previsione) di un numero aleatorio X è definito come la media dei suoi possibili valori pesata con le rispettive probabilità:

$$\mathbb{E}[X] = \sum_i x_i \mathbb{P}\{X = x_i\}$$

se X è discreto e

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

se X è continuo con funzione di densità f.

Se $X = \mathbb{I}_E$ è l'indicatrice di un evento, il suo valore atteso non è altro che la probabilità di E:

$$\begin{aligned} \mathbb{E}[\mathbb{I}_E] &= 0 \times \mathbb{P}(E^c) + 1 \times \mathbb{P}(E) \\ &= \mathbb{P}(E). \end{aligned}$$

In effetti la nozione di valore atteso è una generalizzazione della nozione di probabilità: $\mathbb{E}[X]$ può vedersi come il prezzo equo da pagare, con certezza, per ricevere la quantità aleatoria X (intendendo l'equità nel senso della coerenza).

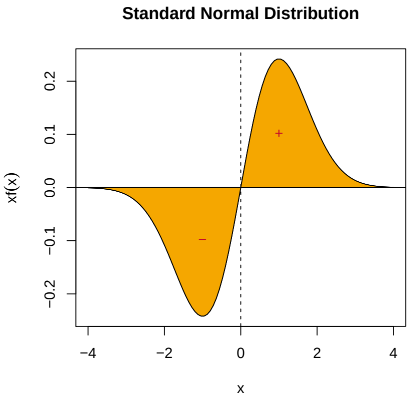
Se X_1 è il punteggio di un dado bilanciato a quattro facce, si ha

$$\begin{aligned}\mathbb{E}[X] &= 1 \times \frac{1}{4} + 2 \times \frac{1}{4} + 3 \times \frac{1}{4} + 4 \times \frac{1}{4} \\ &= \frac{1+2+3+4}{4} = \frac{10}{4} = 2.5;\end{aligned}$$

se A è l'altezza di uno studente preso a caso da una classe di cui sia nota la distribuzione unitaria, il suo valore atteso $\mathbb{E}[A]$ non è altro che la media aritmetica delle modalità osservate.

D'altra parte, se Z segue la distribuzione normale standard, simmetria vuole che il baricentro della sua densità si trovi nell'origine:

$$\mathbb{E}[Z] = 0.$$



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

$$\Rightarrow \int_{-\infty}^{\infty} x f(x) dx = 0$$

Il valore atteso è un operatore lineare:

$$\begin{aligned} \mathbb{E}[X+Y] &= \mathbb{E}[X]+\mathbb{E}[Y], \\ \mathbb{E}[cX] &= c\mathbb{E}[X], \quad c \in \mathbb{R}, \end{aligned}$$

di modo che, per esempio, se S è la somma dei punteggi di due dadi bilanciati a quattro facce, $S = X_1 + X_2$, allora

$$\mathbb{E}[S] \quad = \quad \mathbb{E}[X_1]+\mathbb{E}[X_2] \quad = \quad 2 \times 2.5 \quad = \quad 5.$$

Vale la pena sottolineare che è proprio la linearità a garantire la coerenza di $\mathbb{E}[X]$ come equivalente certo di X al variare del numero aleatorio X.

La varianza di un numero aleatorio X è definita come

$$\mathcal{V}ar(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

ovvero come lo “scarto quadratico atteso dal valore atteso” e può essere calcolata mediante la formula

$$\mathcal{V}ar(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

ovvero come il “valore atteso del quadrato meno il quadrato del valore atteso” (in conseguenza della linearità del valore atteso).

La deviazione standard di un numero aleatorio X è la radice quadrata della sua varianza: $sd(X) = \sqrt{\mathcal{V}ar(X)}$.

Se $X = \mathbb{I}_E$ è l'indicatrice di un evento, si ha

$$\begin{array}{rcccl} \mathbb{E}[\mathbb{I}_E^2] & = & 0^2 \times \mathbb{P}(E^c) + 1^2 \times \mathbb{P}(E) & = & \mathbb{P}(E) \\ \mathcal{V}ar[\mathbb{I}_E] & = & \mathbb{E}[\mathbb{I}_E^2] - \mathbb{E}[\mathbb{I}_E]^2 & = & \mathbb{P}(E) - \mathbb{P}(E)^2 \end{array}$$

di modo che

$$sd[\mathbb{I}_E] = \sqrt{\mathbb{P}(E)(1 - \mathbb{P}(E))};$$

per esempio, se $\mathbb{P}(E) = 1/5$, allora $1 - \mathbb{P}(E) = 4/5$ e

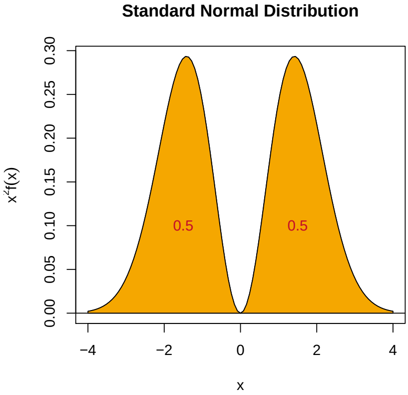
$$sd(\mathbb{I}_E) = \sqrt{\frac{1}{5} \times \frac{4}{5}} = \frac{2}{5}.$$

Se X_1 è il punteggio di un dado bilanciato a quattro facce, si trova

$$\mathcal{V}ar(X_1) = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = \frac{1+4+9+16}{4} - \left(\frac{10}{4}\right)^2$$
$$= \frac{30}{4} - \frac{25}{4} = \frac{5}{4} = 1.25$$

e quindi $sd(X) = \sqrt{5/4} \simeq 1.118$; se A è l'altezza di uno studente preso a caso da una classe di cui sia nota la distribuzione unitaria, la sua varianza e deviazione standard sono quelle delle modalità osservate.

D'altra parte, se Z segue la distribuzione normale standard, allora $\mathcal{V}ar[Z] = \mathbb{E}[Z^2] = 1\ldots$



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

$$\Rightarrow \int_{-\infty}^{\infty} x^2 f(x) dx = 1$$

La deviazione standard di un numero aleatorio X ne misura la variabilità; in particolare il teorema di Chebyshev afferma che

$$\mathbb{P}\{|X - \mathbb{E}[X]| \geq k\, sd(X)\} \leq \frac{1}{k^2}$$

per ogni $k > 0$ (es. $k = 2$ afferma che X dista almeno 2 deviazioni standard da $\mathbb{E}[X]$ con probabilità non superiore a $1/4 = 25\%$).

Per esempio, se Z segue la distribuzione normale standard, si ha

$$\begin{aligned} \mathbb{P}\{|Z| \geq 1\} &\simeq 31.7\% < 100\% \\ \mathbb{P}\{|Z| \geq 2\} &\simeq 4.6\% < 25\% \\ \mathbb{P}\{|Z| \geq 3\} &\simeq 0.3\% < 12\%. \end{aligned}$$

La distribuzione condizionata di un numero aleatorio X, dato un evento E, può descriversi per mezzo della corrispondente funzione di ripartizione condizionata: $\mathbb{P}(X \leq x|E)$, $x \in \mathbb{R}$.

La distribuzione condizionata di X dato E ne determina il valore atteso condizionato $\mathbb{E}[X|E]$: l'equivalente certo di X supponendo il verificarsi di E.

Vediamo subito un'applicazione di questi strumenti nell'ambito delle tavole di mortalità, ma prima chiariamo che (come nel caso dei punteggi ottenuti con due dadi bilanciati a quattro facce)...

…due numeri aleatori X e Y sono stocasticamente indipendenti (o semplicemente indipendenti) secondo la probabilità $\mathbb{P}$ se la distribuzione dell'uno condizionata a un evento espresso per mezzo dell'altro non dipende dall'evento condizionante:

$$\mathbb{P}(X \le x | Y \le y) \quad = \quad \mathbb{P}\{X \le x\}$$

per ogni x e y in $\mathbb{R}$ (è sufficiente considerare eventi di questo tipo).

È immediato riscrivere la condizione di indipendenza come

$$\mathbb{P}(X \leq x \text{ e } Y \leq y) \quad = \quad \mathbb{P}\{X \leq x\} \cdot \mathbb{P}\{Y \leq y\}$$

e, avvalendosi di questa fattorizzazione, estendere la nozione di indipendenza a famiglie di tre o più numeri aleatori.

Se X, Y e Z sono fra loro (a due a due) indipendenti, la varianza della loro somma è pari alla somma delle loro varianze:

$$\mathcal{V}ar(X+Y+Z) = \mathcal{V}ar(X) + \mathcal{V}ar(Y) + \mathcal{V}ar(Z).$$

Un esempio di numero aleatorio con cui ogni mortale si confronta è la durata della vita: per esempio, già ai tempi dell'Antica Roma, il giurista D. Ulpian (?–228) aveva affrontato il problema di stabilire quale fosse il valore attuale di un vitalizio, in funzione dell'età del beneficiario.

Una soluzione matematica di tale problema è basata sull'interpretazione probabilistica delle tavole di mortalità, i cui primordi risalgono alla seconda metà del XVII secolo (lo stesso periodo in cui si gettavano le basi del calcolo delle probabilità nell'ambito dei giochi d'azzardo).

La seguente—famosa—tavola di mortalità (Graunt, 1662) si riferisce agli abitanti di Londra ed è presa da Hald (2003) cui si rinvia per eventuali approfondimenti (sulla tavola e più in generale sulle origini del calcolo delle probabilità).

Age	Survivors	Age	Survivors
0	100	46	10
6	64	56	6
16	40	66	3
26	25	76	1
36	16	86	0

Se T è la durata della mia vita e ho 36 anni, può interessarmi valutare la funzione di sopravvivenza condizionata

$$\mathbb{P}(T \geq t | T \geq 36)$$

per $t > 36$ (per $t \leq 36$ si trova banalmente 100%); per esempio $t = 66$ può darmi la probabilità di andare in pensione...

...la tabella seguente integra la tavola di mortalità di Graunt (1662) con la funzione di sopravvivenza condizionata e quella incondizionata (introducendo una notazione simbolica).

Class (i)	Age (t_i)	Survivors (l_i)	$\mathbb{P}(T \geq t_i)$	$\mathbb{P}(T \geq t_i \mid T \geq 36)$
1	0	100	1.00	1.0000
2	6	64	0.64	1.0000
3	16	40	0.40	1.0000
4	26	25	0.25	1.0000
5	36	16	0.16	1.0000
6	46	10	0.10	0.6250
7	56	6	0.06	0.3750
8	66	3	0.03	0.1875
9	76	1	0.01	0.0625
10	86	0	0.00	0.0000

La mia aspettativa di vita alla nascita $\mathbb{E}[T]$ può scriversi come

$$\mathbb{E}[T] = \frac{\sum_i d_i a_i}{\sum_i d_i}$$

dove: i) a_i è un valore rappresentativo dell'i-esima classe d'età; ii) d_i è il numero di morti nell'i-esima classe d'età.

Analogamente la mia aspettativa di vita attuale è data da

$$\mathbb{E}[T|T \geq 36] - 36 = \frac{\sum_{i \geq k} d_i a_i}{\sum_{i \geq k} d_i} - 36$$

dove k individua la classe che ha 36 come estremo inferiore.

i	t_i	l_i	a_i	d_i	$a_i \times d_i$	$\mathbb{E}[T \mid T \geq t_i]$
1	0	100	3	36	108	18.22
2	6	64	11	24	264	26.78
3	16	40	21	15	315	36.25
4	26	25	31	9	279	45.40
5	36	16	41	6	246	53.50
6	46	10	51	4	204	61.00
7	56	6	61	3	183	67.67
8	66	3	71	2	142	74.33
9	76	1	81	1	81	81.00
10	86	0	Tot.	100	1822	-

Sulla base della precedente tabella, si trova

$$\mathbb{E}[T \,|\, T \geq 36] \quad = \quad \frac{246 + 204 + 183 + 142 + 81}{16} \quad = \quad 53.50$$

di modo che la mia aspettativa di vita attuale vale

$$\mathbb{E}[T \,|\, T \geq 36] - 36 \quad = \quad 53.50 - 36 \quad = \quad 17.50 \quad \text{anni}$$

e quindi, per me, il valore attuale di un vitalizio di 100 euro annui è

$$100 \quad \text{euro / anno} \quad \times \quad 17.50 \quad \text{anni} \quad = \quad 1750 \quad \text{euro,}$$

mentre per un bambino appena nato la tabella fornisce

$$100 \quad \text{euro / anno} \quad \times \quad 18.22 \quad \text{anni} \quad = \quad 1822 \quad \text{euro.}$$

Vi è un metodo alternativo per completare una tavola di mortalità direttamente con le aspettative di vita condizionate: $\mathbb{E}[T|T \geq t_i] - t_i = (L_i + \cdots + L_9)/l_i$, dove $L_i = (t_{i+1} - t_i) \times (l_i + l_{i+1})/2$ sono gli anni vissuti nell'i-esima classe.

i	t_i	l_i	$(l_i + l_{i+1})/2$	L_i	$(L_i + \cdots + L_9)$	$\mathbb{E}[T\|T \geq t_i] - t_i$
1	0	100	82.0	492	1822	18.22
2	6	64	52.0	520	1330	20.78
3	16	40	32.5	325	810	20.25
4	26	25	20.5	205	485	19.40
5	36	16	13.0	130	280	17.50
6	46	10	8.0	80	150	15.00
7	56	6	4.5	45	70	11.67
8	66	3	2.0	20	25	8.33
9	76	1	0.5	5	5	5.00
10	86	0	-	-	-	-

Nel seguito si illustrano alcune delle distribuzioni di probabilità più note e più utili ai fini della statistica inferenziale:

- in primo luogo le più semplici distribuzioni per numeri aleatori discreti (e in particolare la distribuzione binomiale);
- in secondo luogo la distribuzione normale, il cui ruolo è centrale nella statistica inferenziale, avendo cura di discuterne l'uso come approssimazione della distribuzione di un carattere in un campione (o meglio in una popolazione simile al campione);
- in terzo luogo altre distribuzioni continue che giocano un ruolo importante nella statistica inferenziale.

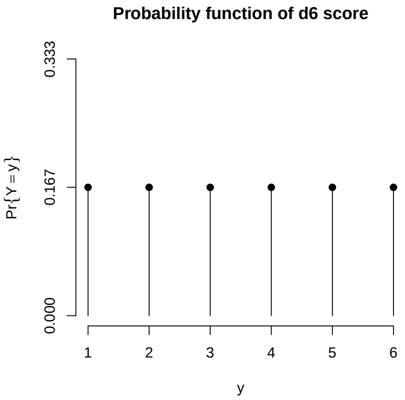
La distribuzione uniforme discreta è quella di un numero aleatorio discreto che può assumere (con probabilità positiva) solo i valori interi compresi in un certo intervallo e per il quale tali valori sono tutti equiprobabili.

La funzione di probabilità di un numero aleatorio Y con distribuzione uniforme discreta è data da

$$\mathbb{P}\{Y = y\} = \frac{1}{s}$$

al variare di y da $z + 1$ (valore minimo) a $z + s$ (valore massimo).

Per esempio ($z = 0$) ha distribuzione uniforme continua il punteggio ottenuto lanciando un dado equilibrato (es. $s = 6$ facce)...



$$\mathbb{E}[Y] = \frac{7}{2} = 3.5$$

$$sd(Y) = \sqrt{\frac{35}{12}} = 1.71$$

Il valore atteso di un numero aleatorio Y con distribuzione uniforme discreta è dato dalla formula

$$\mathbb{E}[Y] = z + \frac{s+1}{2}$$

come si verifica sfruttando la linearità del valore atteso e l'identità $1 + 2 + 3 + \cdots + s = s \cdot (s+1)/2$; analogamente si trova la formula per la varianza

$$\mathcal{V}ar(Y) = \frac{s^2 - 1}{12}$$

prendendo $z = 0$ (senza ledere la generalità) quindi sfruttando l'identità $1 + 4 + 9 + \cdots + s^2 = s \cdot (2s^2 + 3s + 1)/6$ e infine calcolando "il valore atteso del quadrato meno il quadrato del valore atteso".

La distribuzione di Bernoulli è quella di un numero aleatorio discreto che può assumere (con probabilità positiva) solo i valori zero e uno.

La funzione di probabilità di un numero aleatorio Y con distribuzione di Bernoulli è data da

$$\mathbb{P}\{Y = y\} \quad = \quad \psi^y (1 - \psi)^{1-y}$$

al variare di y in $\{0, 1\}$, dove ψ è la probabilità che Y assuma il valore 1.

La distribuzione di Bernoulli è la distribuzione dell'indicatrice dell'evento $\{Y = 1\}$ e pertanto, come già visto, $\mathbb{E}[Y] = \psi$ e $\mathcal{V}ar(Y) = \psi(1 - \psi)$.

La distribuzione binomiale è quella di un numero aleatorio Y che conti i successi in una sequenza di prove ripetute o, più esattamente, di

$$Y = X_1 + X_2 + \cdots + X_n,$$

dove $X_1, X_2, \ldots, X_n$ sono numeri aleatori indipendenti con una comune distribuzione di Bernoulli.

Se ψ è la probabilità di successo (in una qualsiasi prova) la funzione di probabilità di un numero aleatorio Y con distribuzione binomiale è data da

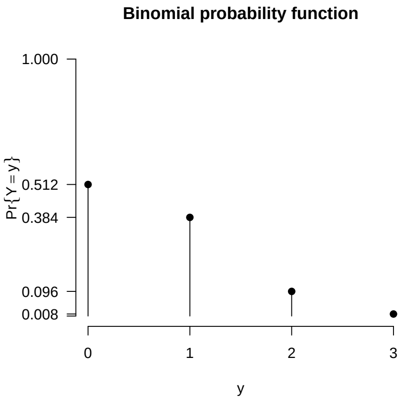
$$\mathbb{P}\{Y = y\} = \frac{n!}{y!(n-y)!}\psi^y(1-\psi)^{n-y}$$

al variare di y in $\{0, 1, 2, \ldots, n\}$, dove $n!$ è il prodotto dei numeri da 1 a n.

Per esempio, se si estraggono <u>con reimmissione</u> tre biglie da un'urna contenente una biglia rossa e quattro gialle, il numero di biglie rosse estratte, Y, avrà una distribuzione binomiale con $n = 3$ e $\psi = 1/5$:

$$\begin{aligned}
\mathbb{P}\{Y = 0\} &= \frac{3!}{0! \times 3!} \times (0.2)^0 \times (0.8)^3 &= 0.512 \\
\mathbb{P}\{Y = 1\} &= \frac{3!}{1! \times 2!} \times (0.2)^1 \times (0.8)^2 &= 0.384 \\
\mathbb{P}\{Y = 2\} &= \frac{3!}{2! \times 1!} \times (0.2)^2 \times (0.8)^1 &= 0.096 \\
\mathbb{P}\{Y = 3\} &= \frac{3!}{3! \times 0!} \times (0.2)^3 \times (0.8)^0 &= 0.008,
\end{aligned}$$

dove convenzionalmente $0! = 1$ e $(0.2)^0 = (0.8)^0 = 1 \ldots$



$$\mathbb{E}[Y] = 3 \times \frac{1}{5} = 0.6$$

$$sd(Y) = \sqrt{3 \times \frac{1}{5} \times \frac{4}{5}} = 0.69$$

Il valore atteso di un numero aleatorio Y con distribuzione binomiale è dato dalla formula

$$\mathbb{E}[Y] = n\psi$$

come si verifica sfruttando la linearità del valore atteso (e la formula per il valore atteso della distribuzione di Bernoulli); analogamente, sfruttando la formula per la varianza della somma di numeri aleatori indipendenti, si trova la varianza di Y

$$\mathcal{V}ar(Y) = n\psi(1-\psi)$$

e quindi (estraendone la radice quadrata) la deviazione standard.

La distribuzione di Poisson (Borra & Di Ciaccio, 2008, p. 234) è quella di un numero aleatorio che conti gli eventi “rari” in un intervallo di tempo fissato, come per esempio i morti in un anno per calcio di cavallo in un corpo di cavalleria dell’esercito prussiano. . .

http://en.wikipedia.org/wiki/Poisson_distribution

. . . può vedersi come limite della distribuzione binomiale per $\psi \to 0$ e $n \to \infty$ mentre il valore atteso $\lambda = n\psi$ resta costante (e risulta essere il valore atteso della distribuzione limite, individuandola univocamente).

La somma di due numeri di Poisson indipendenti, con valori attesi λ e μ, è ancora un numero di Poisson, con valore atteso $\lambda + \mu$.

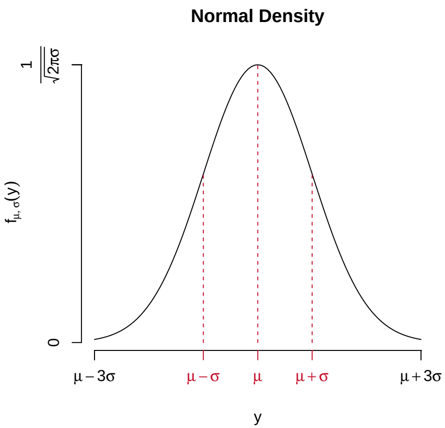
In generale, se Z è normale standard, $\mu \in \mathbb{R}$ e $\sigma > 0$, il numero aleatorio

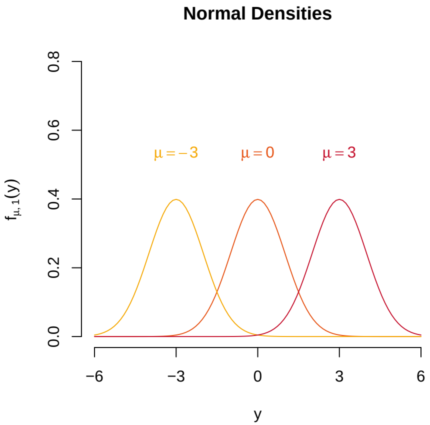
$$Y = \sigma Z + \mu$$

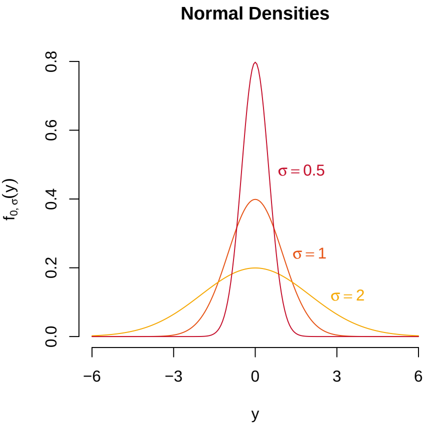
è "normale con media μ e deviazione standard σ", ovvero la sua distribuzione di probabilità è determinata dalla funzione di densità

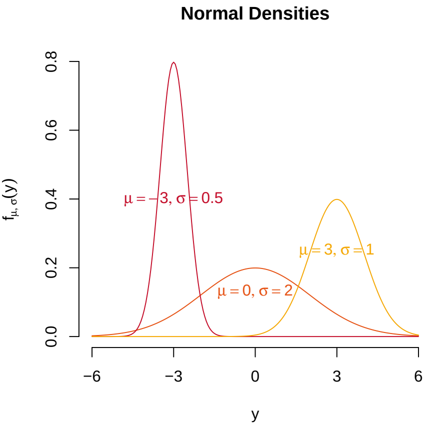
$$f_{\mu,\sigma}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{\sigma^2}}, \quad y \in \mathbb{R}.$$

In questo modo resta definita un'intera famiglia di distribuzioni, ovvero un modello statistico (anche la distribuzione binomiale definisce un modello statistico al variare di ψ in $(0, 1)$).


Normal Density
$\frac{1}{\sqrt{2\pi}\sigma}$
0
$f_{\mu,\sigma}(y)$
$\mu-3\sigma$
$\mu-\sigma$
μ
$\mu+\sigma$
$\mu+3\sigma$
y


Normal Densities
$\mu = -3$
$\mu = 0$
$\mu = 3$
0.0
0.2
0.4
0.6
0.8
$f_{\mu,1}(y)$
−6
−3
0
3
6
y


Normal Densities
$f_{0,\sigma}(y)$
0.0
0.2
0.4
0.6
0.8
$\sigma = 0.5$
$\sigma = 1$
$\sigma = 2$
−6
−3
0
3
6
y


Normal Densities
0.0
0.2
0.4
0.6
0.8
$f_{\mu,\sigma}(y)$
$\mu = -3, \sigma = 0.5$
$\mu = 3, \sigma = 1$
$\mu = 0, \sigma = 2$
−6
−3
0
3
6
y

Se Y è un numero aleatorio normale con media μ e deviazione standard σ, la sua funzione di ripartizione è

$$
\begin{aligned}
F_{\mu,\sigma}(y) &= \mathbb{P}\{Y \leq y\} &&= \mathbb{P}\{\sigma Z + \mu \leq y\} \\
&= \mathbb{P}\{\sigma Z \leq y - \mu\} &&= \mathbb{P}\left\{Z \leq \tfrac{y-\mu}{\sigma}\right\} \\
&= \Phi\left(\tfrac{y-\mu}{\sigma}\right)
\end{aligned}
$$

di modo che, ai fini del calcolo, è sufficiente conoscere Φ (es. Borra & Di Ciaccio, 2004, p. 455) e standardizzare il valore y di interesse:

$$
z = \frac{y-\mu}{\sigma}.
$$

Per esempio, se Y è un numero aleatorio normale con media

$$\mu = 9$$

e deviazione standard

$$\sigma = 2,$$

per fissare le idee la durata in ore della batteria di un telefono cellulare, troveremo

$$\mathbb{P}\{Y \leq 12\} = \mathbb{P}\left\{Z \leq \frac{12-9}{2}\right\} = \mathbb{P}\{Z \leq 1.5\} = 0.9332,$$

come risulta dalla tabella di Borra & Di Ciaccio (2004, p. 455).

Sia ora A l'altezza di uno studente estratto a caso dalla sua classe (in centimetri) e supponiamo di volere valutare le probabilità

- $\mathbb{P}\{174.5 \leq A < 175.5\}$,
- $\mathbb{P}\{A > 195\}$.

Se sono note le altezze di tutti gli studenti della classe, per esempio

169	171	172	173	174	176	176	177	177	178
179	179	179	180	180	180	180	180	181	182
183	184	184	184	185	186	187	188	188	190

in ordine crescente, allora A sarà un numero aleatorio discreto con funzione di probabilità data dalla distribuzione di frequenza relativa delle modalità osservate e $\mathbb{P}\{174.5 \leq A < 175.5\} = \mathbb{P}\{A > 195\} = 0$.

Siamo soddisfatti? Se ci interessa proprio quella classe si, ma se vogliamo riutilizzare la distribuzione di A per un'altra classe simile (o per la generazione di cui quella classe è un campione) forse no...

...a tal fine può convenire (supporre di) conoscere solo la distribuzione di frequenza rispetto a una suddivisione in classi, per esempio

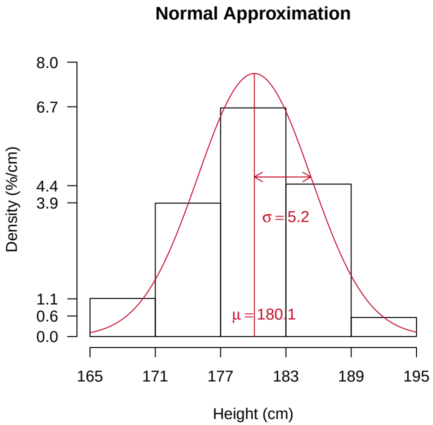
Altezza (cm)	Freq. Ass.	Densità (%/cm)
165 –| 171	2	1.1
171 –| 177	7	3.9
177 –| 183	12	6.7
183 –| 189	8	4.4
189 –| 195	1	0.6
Totale	30	

cosicché si trova

$$\begin{aligned} \mathbb{P}\{174.5 \leq A < 175.5\} &= 1 \times 3.9\% = 3.9\%, \\ \mathbb{P}\{A > 195\} &= 0\%. \end{aligned}$$

La prima risposta è senz'altro più soddisfacente, ai fini di una possibile generalizzazione, ma ha il limite di dipendere dalla suddivisione in classi (e dal fatto che in tale suddivisione gli intervalli siano chiusi a destra); la seconda risposta è ancora insoddisfacente...

...può allora convenire sostituire l'istogramma con un'opportuna curva normale: questa sarà scelta in modo da "adattarsi" ai dati, per esempio prendendo $\mu = m(A) = 5402/30 \simeq 180.1$ e $sd(A) = \sqrt{26.93} \simeq 5.2$, come illustrato dalla figura seguente.


Normal Approximation
8.0
6.7
4.4
3.9
1.1
0.6
0.0
Density (%/cm)
σ = 5.2
μ = 180.1
165
171
177
183
189
195
Height (cm)

L'approssimazione normale ora descritta (che corrisponde ad assumere una distribuzione normale standard per le osservazioni standardizzate) fornisce

$$\mathbb{P}\{174.5 \leq A < 175.5\} = \Phi(z_2) - \Phi(z_1)$$

dove

$$\begin{aligned} z_1 &= \tfrac{174.5-180.1}{5.2} &= -\tfrac{5.6}{5.2} &= -1.08 \\ z_2 &= \tfrac{175.5-180.1}{5.2} &= -\tfrac{4.6}{5.2} &= -0.88 \end{aligned}$$

sono gli estremi dell'intervallo in unità standard; si trova allora

$$\mathbb{P}\{174.5 \leq A < 175.5\} = (1 - 0.8106) - (1 - 0.8599) = 4.93\%,$$

valore che non dipende dalla scelta delle classi (né dalla convenzione di chiuderle a destra).

Analogamente troviamo

$$\begin{aligned} \mathbb{P}\{A > 195\} &= 1 - \Phi\left(\frac{195 - 180.1}{5.2}\right) \\ &= 1 - \Phi(2.87) \\ &= 1 - 0.9979 \\ &= 2.1\%, \end{aligned}$$

valore senza dubbio più soddisfacente di zero...

...tuttavia, evidentemente, l'approssimazione normale è tanto più valida quanto più l'istogramma ha "forma a campana" (simmetrico, 95% delle osservazioni entro ± 2 deviazioni standard dalla media, etc).

Perché proprio la distribuzione normale?

Al di là della facilità d'uso, si possono svolgere due considerazioni:

- da un punto di vista empirico, capita spesso che un istogramma somigli a una curva normale;
- da un punto di vista teorico, vi sono ragioni (teorema del limite centrale) per ritenere che la curva normale sia una buona approssimazione dell'istogramma ogni qual volta la variabilità dei dati (intesa come varianza) possa esprimersi come somma di un numero elevato di contributi indipendenti fra loro, nessuno dei quali prevalente rispetto agli altri.

Queste due considerazioni si rinforzano l'una con l'altra e, in pratica, la distribuzione normale risulta essere la singola distribuzione più importante in statistica.

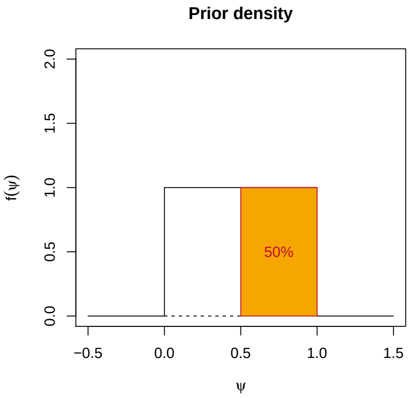
La distribuzione uniforme continua è quella di un numero aleatorio che assuma valori in un intervallo della retta reale in modo che la probabilità di un qualsiasi sottointervallo sia proporzionale alla sua lunghezza.

Un numero uniforme continuo ammette densità

$$f(y) = \frac{1}{b-a}$$

per $a \leq y \leq b$ e zero per valori di y esterni all'intervallo (a, b).

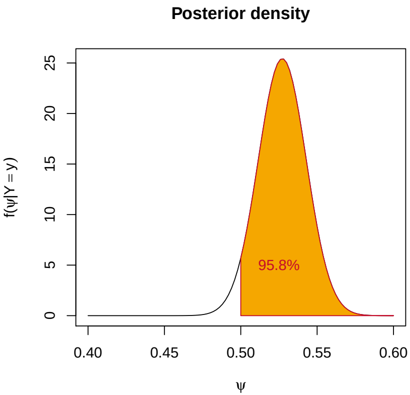
Per esempio una distribuzione uniforme continua potrebbe esprimere la nostra incertezza iniziale sul parametro ψ di un modello binomiale (es. proporzione di individui favorevoli a una certa proposta)...



$$a = 0 \quad \& \quad b = 1$$

$$\mathbb{E}[Y] = \frac{a+b}{2} = 0.5$$

$$sd(Y) = \frac{(b-a)}{2\sqrt{3}} = 0.29$$



Intervistati:

$$n = 1011$$

Favorevoli:

$$y = 533$$

. . . dopo di che (sulla base di una variante del teorema di Bayes) la nostra incertezza finale su ψ (proporzione incognita di individui favorevoli alla proposta) alla luce di un campione di intervistati (dato binomiale) sarebbe espressa da una distribuzione beta (Borra & Di Ciaccio, 2008, p. 248).

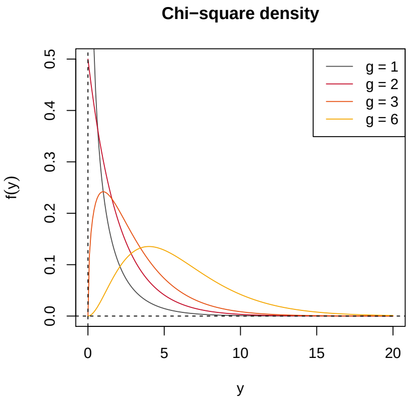
Sulla base di questa analisi bayesiana (qui appena accennata) concluderemmo che si tratta di una maggioranza con probabilità pari al 95.8% (contro il 50% iniziale).

La distribuzione chi-quadrato è quella di un numero aleatorio Y che sia somma di quadrati di numeri normali standard indipendenti:

$$Y = Z_1^2 + Z_2^2 + \cdots + Z_g^2,$$

dove $Z_1, Z_2, \ldots Z_g$ sono numeri aleatori indipendenti con distribuzione normale standard e g sono i gradi di libertà della distribuzione χ^2.

La figura seguente riporta i grafici di alcune densità chi-quadrato con diversi gradi di libertà e Borra & Di Ciaccio (2008, p. 494) tabulano alcuni percentili (estremi) per distribuzioni chi-quadrato con gradi di libertà da 1 a 100; per valori più elevati di g si può usare un'approssimazione normale, tenendo conto che $\mathbb{E}[Y] = g$ e $\mathcal{V}ar(Y) = 2g$.



Il rapporto tra un numero normale standard e la radice quadrata di un numero χ^2 da esso indipendente, diviso per i suoi gradi di libertà, ha una distribuzione t di Student (Borra & Di Ciaccio, 2008, p. 245).

Il rapporto tra due numeri χ^2 indipendenti, ognuno diviso per i propri gradi di libertà, ha una distribuzione *F* di Fisher-Snedecor (Borra & Di Ciaccio, 2008, p. 247)

La distribuzione esponenziale (Borra & Di Ciaccio, 2008, p. 248) è quella di un numero aleatorio continuo, Y, che esprima la durata di un qualcosa che non invecchi (es. una lampadina a incandescenza):

$$\mathbb{P}(Y > y_0 + y | Y > y_0) = \mathbb{P}(Y > y);$$

per esempio ($y_0 = 50$, $y = 10$) la probabilità che una lampadina (con durata esponenziale) duri altre 10 ore, dopo essere stata accesa per 50 ore, sarà pari alla probabilità che tale lampadina, appena accesa, duri 10 ore.

Vale la pena osservare che la distribuzione esponenziale ha funzione di densità monotona decrescente e che per tale distribuzione si ha

$$\mathbb{E}[Y] = sd(Y).$$

BORRA, S. & DI CIACCIO, A. (2008).
Statistica: Metodologie per le Scienze Economiche e Sociali (Seconda Edizione).
McGraw-Hill, Milano.

GRAUNT, J. (1662).
Natural and Political Observations made upon the Bills of Mortality.
Martyn, London.

HALD, A. (2003).
History of Probability and Statistics and Their Applications before 1750.
Wiley, Hoboken.