5 Adattamento di una distribuzione teorica ad una distribuzione di frequenza empirica

Osservata una distribuzione di frequenza empirica, uno degli obiettivi più importanti dell'analisi statistica dei dati è quello di individuare una opportuna legge o funzione analitica, che riesca efficacemente a descrivere il carattere rilevato. Tale funzione, se presenta determinate proprietà, prende il nome di "*distribuzione teorica*".

Le caratteristiche di una distribuzione teorica possono essere estese al fenomeno oggetto di studio, il cui comportamento non è noto, al fine di rendere più agevole l'approccio matematico.

Le distribuzioni teoriche sono rappresentate da modelli probabilistici, che descrivono l'andamento di particolari variabili, dette *variabili casuali*.

5.1 Cenni di calcolo delle probabilita'

Si definisce *evento casuale o aleatorio* il risultato di un esperimento, definito aleatorio perché non si può prevedere.

Definizione classica di probabilità

Secondo l'approccio classico, la probabilità di un evento A è il rapporto tra il numero dei casi favorevoli al verificarsi di A e il numero dei casi possibili, purché questi ultimi siano equiprobabili.

esempi:

a) Nell'esperimento "lancio di una moneta", si calcoli la probabilità dell'evento T "esce testa".

Casi favorevoli: T

Casi possibili: T, C (esce croce)

T e C hanno la stessa probabilità di uscire se la moneta non è truccata:

P(T)= P(C)=1/2.

b) Nell’esperimento “lancio di un dado”, si calcoli la probabilità dell’evento E “esce numero pari”.

Casi favorevoli: 2, 4, 6

Casi possibili: 1, 2, 3, 4, 5, 6.

P(E)=3/6.

Definizione frequentista di probabilità

Secondo l’approccio frequentista, la probabilità di un evento è il limite ($n \to \infty$) della frequenza relativa dei successi.

Definizione soggettivista di probabilità

Secondo l’approccio soggettivista, la probabilità è il risultato di una valutazione soggettiva da parte di un individuo. Per es., l’ideatore di un giocattolo può assegnare alla probabilità che il giocattolo abbia successo sul mercato un valore diverso rispetto al responsabile marketing della società che vende il giocattolo stesso. L’assegnazione di una probabilità soggettiva ad un evento tiene conto:

- dell’esperienza passata dell’individuo;
- della sua opinione personale;
- dell’analisi del particolare contesto di riferimento.

L’approccio soggettivista è particolarmente utile quando la probabilità di un evento non può essere determinata empiricamente.

Esiste una corrispondenza biunivoca tra eventi ed insiemi, per cui fra gli eventi è possibile effettuare tutte le operazioni lecite fra gli insiemi, per esempio l’$\cup$ e l’$\cap$; l’unione corrisponde all’o disgiuntivo, mentre l’intersezione corrisponde all’e congiunzione.

Definizione assiomatica di probabilità

La probabilità si può definire in via assiomatica nel seguente modo:

La probabilità di un evento E è quel numero reale p tale che:

1) p=P(E)$\geq$0;

2) P(Ω)=1;

3) $P(E_1 \cup E_2)=P(E_1)+P(E_2)$ se $E_1 \cap E_2=\emptyset$, cioè se E_1 ed E_2 sono "incompatibili" (*teorema delle probabilita' totali per eventi incompatibili*)

Esempio di eventi incompatibili

Nell'esperimento "lancio di un dado" gli eventi:

E_1=esce numero pari;

E_2=esce numero dispari

sono incompatibili, perché il verificarsi dell'uno esclude il verificarsi dell'altro, ossia E_1 ed E_2 non possono verificarsi entrambi contemporaneamente.

La probabilità, dunque, è un numero reale che varia tra 0 e 1; associamo il valore 0 ad un evento che non ha nessuna probabilità di verificarsi (**evento impossibile**: Φ) e il valore 1 ad un evento che si verificherà sicuramente (**evento certo:** Ω).

Ω è lo *spazio dei risultati (elementari)*, ovvero è l'insieme di tutti i possibili risultati incompatibili connessi ad un esperimento. Si consideri, ad esempio, il lancio di due dadi. Lo spazio dei risultati è:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,6)	(6,6)

Come si è visto, l'insieme vuoto $\emptyset$ in Calcolo delle probabilità viene definito "*evento impossibile*". Dimostriamo che:

P($\emptyset$)=0.

Dimostrazione:

$P(\Omega)=P(\Omega\cup\varnothing)=P(\Omega)+P(\varnothing)=1$, dunque $P(\varnothing)=0$.

Se $E_1\cap E_2\neq\varnothing$, allora $P(E_1\cup E_2)=P(E_1)+P(E_2)-P(E_1\cap E_2)$.

Se $E_1\cap E_2=\varnothing$, allora $P(E_1\cap E_2)= P(\varnothing)=0$, da cui segue il 3° assioma.

Vediamo adesso come si modifica il calcolo di una probabilità quando si dispone di un'informazione a priori sugli eventi coinvolti.

Si definisce **probabilità condizionata** la probabilità che si verifichi un evento E_1 sapendo che l'evento E_2 si è già verificato:

$$P(E_1| E_2) = P(E_1 \cap E_2) / P(E_2)$$

Analogamente

$$P(E_2| E_1) = P(E_2 \cap E_1) / P(E_1),$$

da cui segue la *legge delle probabilità composte*:

$P(E_1 \cap E_2) = P(E_2 \cap E_1) = P(E_1| E_2) \cdot P(E_2) = P(E_2| E_1) \cdot P(E_1)$.

Quando il verificarsi di un evento non influenza la probabilità che se ne verifichi un altro, si dice che i due eventi sono **indipendenti**:

$$P(E_1| E_2) = P(E_1),$$

da cui segue il *Teorema delle probabilita' composte per eventi indipendenti*:

Due eventi sono indipendenti se e solo se $P(E_1\cap E_2)= P(E_1)\cdot P(E_2)$.

Esempio:

Si abbia un'urna con 10 palline, di cui 7 bianche (B) e 3 nere (N). Si considerino gli eventi:

E_1=esce pallina bianca alla 1° estrazione;

E_2=esce pallina bianca alla 2° estrazione.

Si vuole calcolare la probabilità che esca pallina B alla 1° ed alla 2° estrazione, ovvero $P(E_1 \cap E_2)$. Si ha:

$P(E_1)=7/10$

$P(E_2|E_1)= P(E_2)=7/10$ se l'estrazione è effettuata con ricollocamento (R),

$P(E_2|E_1)= 6/9$ se l'estrazione è effettuata senza ricollocamento.

Allora, $P(E_1 \cap E_2) = P(E_1) \cdot P(E_2) = (7/10) \cdot (7/10)$ solo se l'estrazione è effettuata con R; in tal caso, i due eventi sono indipendenti.

Se l'estrazione è effettuata senza R, $P(E_1 \cap E_2) = P(E_1) \cdot P(E_2 | E_1) = (7/10) \cdot (6/9)$.

Se, invece, l'estrazione è effettuata senza R ed è E_1=esce pallina nera alla 1° estrazione, allora $P(E_1 \cap E_2) = P(E_1) \cdot P(E_2 | E_1) = (3/10) \cdot (7/9)$.

VARIABILE CASUALE

Consideriamo l'esperimento "lancio di una moneta 2 volte"; i risultati possibili sono: CC, CT, TC, TT.

Si può definire la funzione "n. di volte che si presenta T"; tale funzione assume valore 0, 1, 2 e viene definita *variabile casuale (v.c.)*.

Ai singoli valori 0, 1, 2 si possono associare le probabilità 1/4, 2/4, 1/4:

x_i	p_i
0	1/4
1	2/4
2	1/4
totale	1

La v.c. considerata è una v.c. discreta, che assume un numero finito di valori, ma esistono anche v.c. discrete che assumono un'infinità numerabile di valori e v.c. continue.

VARIABILE CASUALE DISCRETA

Una v.c. discreta è una funzione che può assumere un numero finito o un'infinità numerabile di valori, ai quali sono associate probabilità note, la cui somma è uguale a 1.

La distribuzione di probabilità associata alla v.c. discreta è dunque una funzione p(x), che gode delle seguenti proprietà:

- $p(x) \geq 0$;
- $\Sigma_x p(x)=1$.

VARIABILE CASUALE CONTINUA

Una v.c. continua è una funzione che può assumere infiniti valori all'interno di un intervallo, limitato o illimitato.

La distribuzione di probabilità associata alla v.c. continua prende il nome di "funzione di densità di probabilità" e gode delle seguenti proprietà:

- $f(x) \geq 0$;
- $\int_R f(x)dx=1$.

MOMENTO TEORICO DI ORDINE *r* E ORIGINE *m*

Per una v.c. discreta il *momento teorico di ordine r e origine m* è definito dalla seguente espressione:

$$\mu_{m,r} = \sum_x (x-m)^r p(x)$$

mentre per una v.c. continua, è definito dall'espressione:

$$\mu_{m,r} = \int_{-\infty}^{+\infty} (x-m)^r f(x)dx .$$

Un particolare momento è il valore atteso $E(X) = \mu_{0,1}$.

MOMENTO TEORICO CENTRATO DI ORDINE r

Si ottiene quando l'origine *m* è uguale al valore atteso $E(X)$:

$$\mu_r = \sum_x [x - E(X)]^r p(x) \qquad \text{per una v.c. discreta}$$

$$\mu_r = \int_{-\infty}^{+\infty} [x - E(X)]^r f(x)dx \qquad \text{per una v.c. continua}$$

Un particolare momento centrato è la varianza $Var(X) = \mu_2$.

SIMMETRIA

Una v.c. discreta $X=x_1, x_2,\dots,x_n$ ha distribuzione di probabilità simmetrica quando:

$p(x_1)=p(x_n)$

$p(x_2)=p(x_{n-1})$

$p(x_3)=p(x_{n-2})$

e così via.

Una v.c. continua X ha funzione di densità simmetrica quando, comunque preso $h>0$,

$f(x_0+h)=f(x_0-h)$,

essendo $x=x_0$ l'asse di simmetria. Spesso $x_0=M_0$, punto in cui corrisponde il massimo di $f(x)$.

La simmetria è una proprietà molto importante per una distribuzione.

Per avere una misura dell'asimmetria di una distribuzione, si può utilizzare l'indice β_1, dato dal rapporto:

$$\beta_1 = \frac{\mu_3}{\sigma^3}$$

Se la distribuzione è simmetrica, i momenti di grado dispari, come μ_3, sono nulli, di conseguenza, per una distribuzione simmetrica è $\beta_1 = 0$.

CURTOSI

Un'altra caratteristica importante della forma di una distribuzione è la curtosi, di cui si parlerà più avanti (cfr.par. 5.4). La curtosi può essere misurata dal rapporto:

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}.$$

5.2 La distribuzione binomiale

Supponiamo di avere un'urna con N palline, di cui B bianche ed $N-B=\overline{B}$ di un altro colore.

La probabilità di estrarre una pallina *B* è :

P(B)=B/N=p.

La probabilità di estrarre una pallina di altro colore è:

$$P\left(\overline{B}\right)=\frac{\overline{B}}{N}=q=\frac{N-B}{N}=1-\frac{B}{N}=1-p\ .$$

Esperimenti di questo tipo, i cui possibili risultati sono costituiti da due eventi, un "successo" (esce pallina *B*) e un "insuccesso" (esce pallina $\overline{B}$), con probabilità, rispettivamente, *p* e *q=1-p*, vanno sotto il nome di "ESPERIMENTI BERNOULLIANI".

Supponiamo di effettuare *n* estrazioni con R e di essere interessati all'evento "esce pallina *B* alla 1° ed alla 2° estrazione". Poiché le prove sono indipendenti, la probabilità di tale evento, per il "teorema delle probabilità composte per eventi indipendenti", è:

$$pp\underbrace{qq......q}_{n-2}$$

Se vogliamo calcolare la probabilità di estrarre pallina bianca alla 1°, alla 2°,….., alla *x-ma* estrazione, tale probabilità sarà allora:

$$\underbrace{pp...p}_{x}\underbrace{qq...q}_{n-x}$$

Se non siamo interessati all'ordine, ossia se vogliamo calcolare la probabilità che, su *n* estrazioni, esca genericamente "*x* volte" pallina B, dovremmo tener conto di tutte le possibili sequenze:

$$\underbrace{pp...p}_{x}\underbrace{qq...q}_{n-x}$$

$$\underbrace{qq...q}_{n-x}\underbrace{pp...p}_{x}$$

$$pp\underbrace{qq...q}_{n-x}p...p$$

……………

Tutte le possibili sequenze sono $\binom{n}{x}$, dunque, per il "teorema delle probabilità totali per eventi incompatibili", la probabilità cercata è:

$$P(X = x) = p(x) = \binom{n}{x} p^x q^{n-x}$$

dove $\binom{n}{x} = \dfrac{n!}{x!(n-x)!}$.

$p(x)$ è una funzione di distribuzione di probabilità chiamata "BINOMIALE".

Essa consente di calcolare la probabilità di ottenere *x* successi in *n* prove.

La v.c. cui è associata prende il nome di v.c. binomiale; la v.c. binomiale è discreta e può assumere un numero finito di valori: *x=0,1,2,…,n*.

La funzione binomiale è una funzione di distribuzione di probabilità, infatti:

- $p(x) = \binom{n}{x} p^x q^{n-x} \geq 0$, poiché è il prodotto di numeri positivi (*n*, *x*, *p* e *q* sono maggiori o al più uguali a 0);
- $\sum_{x=0}^{n} p(x) = (p+q)^n = (p+1-p)^n = 1$.

La distribuzione binomiale è così chiamata perché rappresenta il generico elemento dello sviluppo del binomio di Newton:

$$(p+q)^n = \binom{n}{0} p^0 q^n + \binom{n}{1} p^1 q^{n-1} + \ldots + \binom{n}{x} p^x q^{n-x} + \ldots + \binom{n}{n} p^n q^0 = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x}$$

.

La distribuzione binomiale è caratterizzata da due parametri: *n* e *p*.

Si dimostra che i momenti teorici della distribuzione binomiale sono:

$$\mu_{0,1} = E(X) = \sum_{x=0}^{n} x p(x) = np$$

$$\mu_2 = Var(X) = \sum_{x=0}^{n} [x - E(X)]^2 \, p(x) = npq$$

$$\mu_3 = \sum_{x=0}^{n} [x - E(X)]^3 \, p(x) = npq(q - p).$$

Di conseguenza è:

$$\beta_1 = \frac{\mu_3}{\sigma^3} = \frac{npq(q-p)}{(npq)^{3/2}} = \frac{(q-p)}{(npq)^{1/2}}$$

Se *p*=*q*=1/2, sarà (*q*-*p*)=0, dunque $\beta_1 = 0$: la distribuzione binomiale è simmetrica. Ciò si verifica anche quando n$\rightarrow\infty$, poiché in tal caso la binomiale tende alla distribuzione di Gauss (cfr.par. 5.4).

5.3 La distribuzione di Poisson

Se $n\rightarrow\infty$ e con la stessa velocità $p\rightarrow 0$, cioè se $n\rightarrow\infty$ in modo che $np=\lambda$ resti costante, la distribuzione binomiale tende ad una distribuzione limite che va sotto il nome di “distribuzione di Poisson”:

$$P(X = x) = p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x=0, 1, 2,\ldots$$

esempio

Si abbia un’urna con *N* palline, di cui *B* bianche.

Effettuate *n* estrazioni con R, la distribuzione binomiale consente di calcolare la probabilità che escano *x* palline B su *n*.

Se, però, la percentuale di palline B è molto bassa ($p\rightarrow 0$), è necessario aumentare il n. delle estrazioni (n$\rightarrow\infty$) affinché, in media, si possa osservare sempre lo stesso n. di palline B, affinché cioè $np=\lambda$ resti costante.

La distribuzione di POISSON, poiché $p\rightarrow 0$, viene definita anche distribuzione degli EVENTI RARI (esempi: n. di morti, n. guasti, ecc…).

La v.c. di Poisson è una v.c. discreta, che assume un'infinità numerabile di valori; infatti, poiché $n \to \infty$, $x = 0, 1, 2, 3, \ldots$

La distribuzione di Poisson è caratterizzata da un solo parametro: λ.

Essa è una funzione di probabilità:

1) $p(x) \geq 0$, perché quoziente di quantità positive: $x \geq 0$, $\lambda = np > 0$, $e > 0$;

2) $\sum_{x=0}^{\infty} p(x) = 1$, poiché:

$$\sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1.$$

Si dimostra che i momenti teorici della Poisson sono:

$$\mu_{0,1} = E(X) = \sum_{x=0}^{\infty} x p(x) = \lambda$$

$$\mu_2 = Var(X) = \sum_{x=0}^{\infty} [x - E(X)]^2 p(x) = \lambda$$

$$\mu_3 = \sum_{x=0}^{\infty} [x - E(X)]^3 p(x) = \lambda .$$

Dunque $E(X) = Var(X) = \lambda$. L'uguaglianza tra $E(X)$ e $Var(X)$ è una peculiarità della Poisson, mentre per la binomiale $E(X) > Var(X)$.

Inoltre, essendo:

$$\beta_1 = \frac{\mu_3}{\sigma^3} = \frac{\lambda}{(\lambda)^{3/2}} = \frac{1}{(\lambda)^{1/2}} > 0,$$

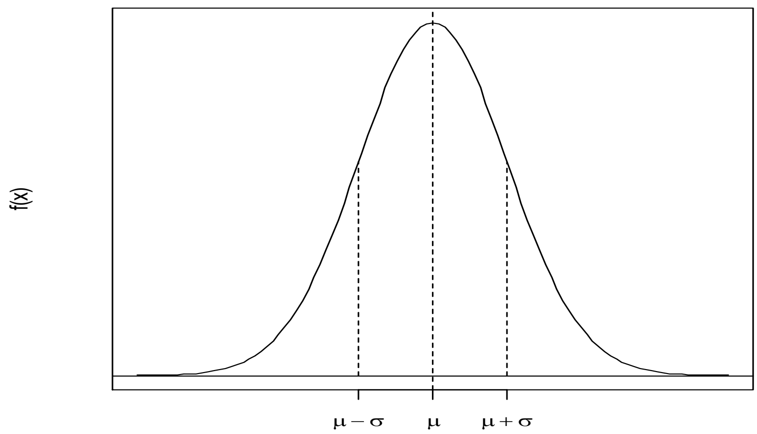
la distribuzione di Poisson è sempre asimmetrica positivamente.

5.4 La distribuzione normale o di Gauss

Quando $n \to \infty$, ma p assume qualsiasi valore compreso tra 0 e 1, la distribuzione binomiale tende ad un'altra distribuzione limite. Si tratta, però, questa volta di un modello teorico continuo, che prende il nome di "distribuzione normale":

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad \text{con } -\infty < \mu < +\infty \text{ e } \sigma > 0 .$$

Questa funzione, nota anche come distribuzione di Gauss, è definita su un intervallo illimitato: $-\infty < x < +\infty$, è simmetrica di forma campanulare e asintotica rispetto all'asse X. Presenta un punto di massimo in corrispondenza di x=μ, e due punti di flesso in corrispondenza di *(μ-σ) e (μ+σ)*:



La curva sopra descritta viene considerata la legge di distribuzione per eccellenza degli errori accidentali. In realtà, gli errori accidentali, pur avendo una distribuzione di probabilità simmetrica, non sempre seguono la legge di Gauss (cfr.par. 3.3 e 4.1.1).

La legge di Gauss è una "funzione di densità di probabilità". Si dimostra, infatti, che sono verificate le due condizioni:

- $f(x) \geq 0$

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

Si dimostra, inoltre, che i due parametri che caratterizzano la distribuzione, μ e σ^2, sono proprio il valore atteso e la varianza.

Infatti, i momenti teorici della *f(x)* sono dati dalle seguenti espressioni:

$$\mu_{0,1} = \int_{-\infty}^{+\infty} x f(x)dx = \mu$$

$$\mu_2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx = \sigma^2$$

$$\mu_3 = \int_{-\infty}^{+\infty} (x-\mu)^3 f(x)dx = 0$$

$$\mu_4 = \int_{-\infty}^{+\infty} (x-\mu)^4 f(x)dx = 3\sigma^4$$

μ_3 e μ_4 consentono di calcolare, rispettivamente, l'indice di asimmetria β_1 e l'indice di curtosi β_2.

Per una distribuzione normale risulta

$$\beta_1 = \frac{\mu_3}{\sigma^3} = \frac{0}{\sigma^3} = 0$$
$$\beta_2 = \frac{\mu_4}{\sigma^4} = \frac{3\sigma^4}{\sigma^4} = 3.$$

La distribuzione normale viene definita "mesocurtica". Altre curve simmetriche, più appuntite o più appiattite rispetto alla normale, vengono definite, rispettivamente "leptocurtiche" ($\beta_2>3$) e "platicurtiche" ($\beta_2<3$). La distribuzione di Laplace è una distribuzione leptocurtica ($\beta_2=6$), mentre la distribuzione uniforme è platicurtica ($\beta_2=1,8$).

Per una variabile casuale continua non possiamo calcolare probabilità puntuali, o meglio, la probabilità che una v.c. continua X assuma esattamente un valore x è zero:

$P(X=x)=0$.

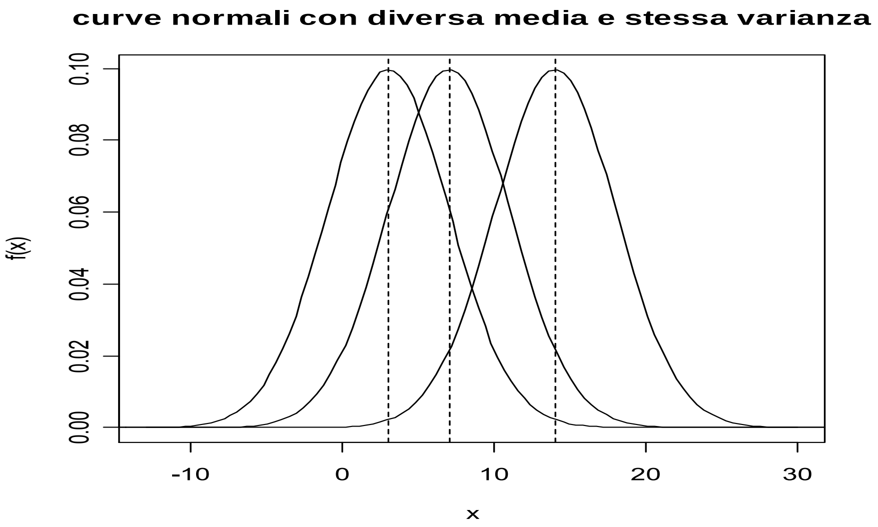
Possiamo, però, calcolare la probabilità che una v.c. continua assuma valori in un determinato intervallo, sia esso limitato o illimitato. Calcolare una probabilità di questo tipo equivale a calcolare un'area, ad esempio:

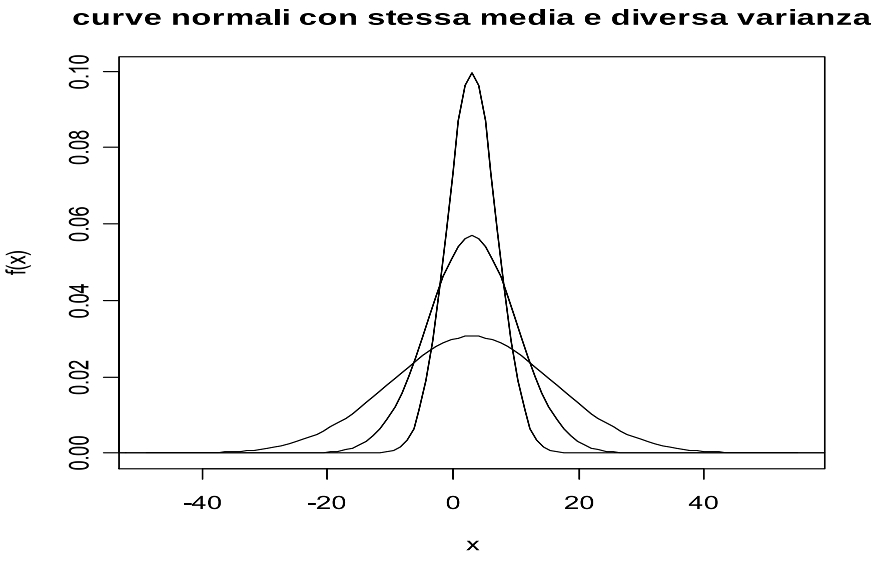
$$P(x_i < X < x_{i+1}) = \int_{x_i}^{x_{i+1}} f(x)dx = \int_{-\infty}^{x_{i+1}} f(x)dx - \int_{-\infty}^{x_i} f(x)dx = P(X < x_{i+1}) - P(X < x_i).$$

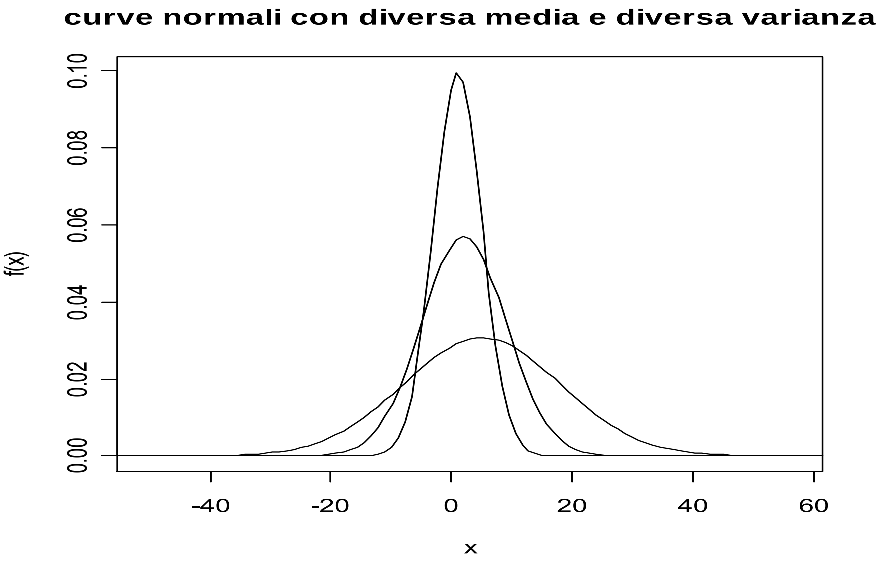
Per cui, se $x_i=x_{i+1}$, allora $P(x_i < X < x_{i+1}) = \int_{x_i}^{x_i} f(x)dx = 0$.

L'integrale $\int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$ $=P(X < x)=F(x)$, noto come "funzione di ripartizione di *X*, non è risolvibile in forma chiusa, ma tramite procedimenti di Analisi numerica.

Calcolare questo integrale per ogni curva normale sarebbe stato impensabile, poiché i valori che μ e σ^2 possono assumere sono infiniti:







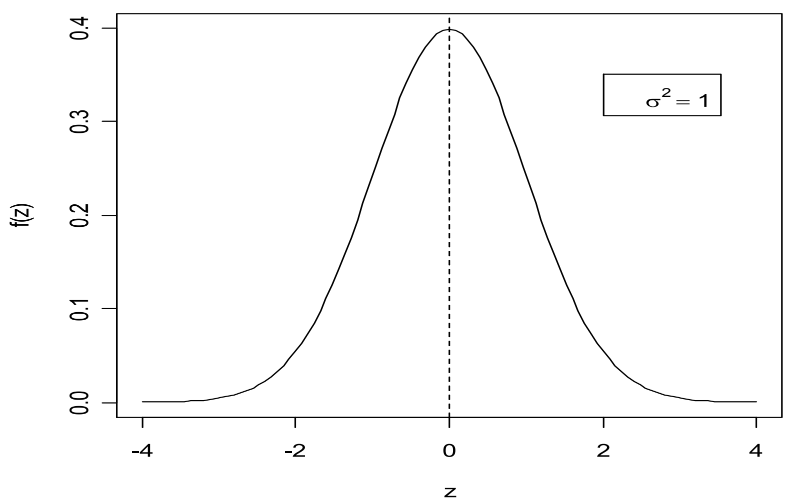
E’ dunque risultato conveniente considerare la trasformata *Z* di *X*:

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

Z è una particolare v.c. normale, chiamata “*v.c. normale standardizzata*”, la cui densità è:

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

che ha la peculiarità di avere $\mu=0$ e $\sigma^2=1$:



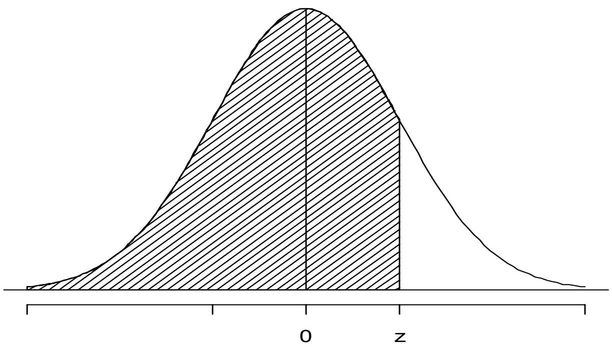
Sono stati calcolati, allora, gli integrali del tipo:

$$\int_{-\infty}^{z} f(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = P(Z < z) = F(z).$$

Tali integrali sono stati tabulati in appositi prontuari, chiamati "*prontuari delle probabilità integrali della curva normale standardizzata*".

Calcolare l'integrale

$$\int_{-\infty}^{z} f(t)dt = P(Z < z) = F(z)$$



equivale a calcolare l'integrale

L'integrale $\int_{-\infty}^{x} f(t)dt = P(X < x) = F(x).$

Conviene quindi eseguire la trasformazione

$$z_i = \frac{x_i - \mu}{\sigma}$$

e trasformare gli intervalli x_i-x_{i+1} in z_i-z_{i+1}.

Calcolare sul prontuario le probabilità $P(Z<z_i)$ *e* $P(Z<z_{i+1})$, *dunque la probabilità* $P(z_i<Z<z_{i+1})=P(Z<z_{i+1})-P(Z<z_i)$, equivale a calcolare la probabilità $P(x_i<X<x_{i+1})$.

Particolare interesse assumono le probabilità dei valori compresi in intervalli simmetrici intorno alla media, di ampiezza pari ad un multiplo dello scarto quadratico medio:

$P(\mu-\sigma<X<\mu+\sigma)=P(-1<Z<+1)=0{,}68268=68{,}3\%$

$P(\mu-2\sigma<X<\mu+2\sigma)=P(-2<Z<+2)=0{,}95450=95{,}4\%$

$P(\mu-3\sigma<X<\mu+3\sigma)=P(-3<Z<+3)=0{,}99730=99{,}7\%$.

Altrettanto importanti risultano le seguenti probabilità:

$P(\mu-1{,}96\sigma<X<\mu+1{,}96\sigma)=P(-1{,}96<Z<+1{,}96)=0{,}95$

$P(\mu-2{,}58\sigma<X<\mu+2{,}58\sigma)=P(-2{,}58<Z<+2{,}58)=0{,}99$

$P(\mu-3{,}29\sigma<X<\mu+3{,}29\sigma)=P(-3{,}29<Z<+3{,}29)=0{,}999$.

A scopo esemplificativo, servendoci del prontuario delle probabilità integrali di una normale standardizzata, proviamo a calcolare l'ultima:

$P(\mu-3{,}29\sigma<X<\mu+3{,}29\sigma)=P(-3{,}29<Z<+3{,}29)=P(Z<3{,}29)-P(Z<-3{,}29)=F(3{,}29)-F(-3{,}29)= F(3{,}29)-[1-F(3{,}29)]=2\cdot F(3{,}29) -1=2\cdot 0{,}9995 -1=0{,}999$.

5.5 Adattamento di una distribuzione teorica ad una distribuzione empirica

Osservata una distribuzione di frequenza empirica, vediamo adesso quali sono i criteri che ci portano ad adattare un determinato modello teorico piuttosto che un altro.

In linea di massima, la nostra scelta non potrà che ricadere sui tre modelli analizzati: quello binomiale, quello di Poisson e quello di Gauss.

Consideriamo il seguente esempio. Un collettivo di 80 studenti è stato sottoposto ad un test attitudinale per l'ammissione ad un corso di matematica. Viene riportata la distribuzione di frequenze del *numero di errori commessi* su un totale di 10 domande:

n. di errori commessi x_i	frequenze assolute osservate n_i
0	1
1	2
2	6
3	9
4	14
5	22
6	12
7	7
8	5
9	1
10	1
Totale	80

La variabile statistica osservata "*n. di errori commessi*" è una variabile quantitativa discreta, per cui la nostra preferenza verte su una delle due v.c. discrete studiate: la v.c. binomiale o la v.c. di Poisson. Fra la v.c. binomiale e la v.c. di Poisson scegliamo la v.c. binomiale, poiché la media empirica risulta maggiore della varianza empirica e poiché la variabile *X* non sembra descrivere un evento raro:

n. di errori commessi x_i	frequenze assolute osservate n_i	$x_i n_i$	x_i^2	$x_i^2 n_i$
0	1	0	0	0
1	2	2	1	2
2	6	12	4	24
3	9	27	9	81
4	14	56	16	224
5	22	110	25	550
6	12	72	36	432
7	7	49	49	343
8	5	40	64	320
9	1	9	81	81
10	1	10	100	100
totale	80	387		2157

$$M = \frac{\sum_{i=1}^{n} x_i n_i}{N} = 4{,}84 \qquad \sigma^2 = \frac{\sum_{i=1}^{n} x_i^2 n_i}{N} - M^2 = \frac{2157}{80} - (4{,}84)^2 = 3{,}56 .$$

Scelto il modello binomiale, essendo *n=10* noto, dobbiamo stimarne solo il parametro *p*. A tal fine potremmo utilizzare il **metodo dei momenti**, che consiste nell'uguagliare momenti empirici e momenti teorici. Supponiamo, solo a scopo esemplificativo, che anche *n* non sia noto; bisogna dunque risolvere il sistema:

$$\begin{cases} M = np \\ \sigma^2 = npq \end{cases}$$

Dal sistema, sostituendo nella seconda equazione *M* ad *np*, si ricava che:

$$\hat{q} = \frac{\sigma^2}{M} = \frac{3{,}56}{4{,}84} = 0{,}74 \qquad \text{da cui } \hat{p} = 1 - \hat{q} = 0{,}26 .$$

Inoltre, dalla prima equazione è:

$$\hat{n} = \frac{M}{\hat{p}} = \frac{4{,}84}{0{,}26} \cong 19 .$$

Possiamo adesso calcolare le probabilità teoriche, al variare di x:

$$p_i = p(x) = \binom{\hat{n}}{x} \hat{p}^x \hat{q}^{\hat{n}-x} = \binom{19}{x} (0{,}26)^x (0{,}74)^{19-x}.$$

Moltiplichiamo, quindi, le probabilità teoriche per il totale delle osservazioni, in modo tale da ottenere le "frequenze teoriche" n_i^*, che devono essere poste a confronto con le frequenze empiriche n_i; più le frequenze teoriche si avvicineranno alle frequenze empiriche, migliore sarà l'adattamento del nostro modello scelto ai dati osservati.

Pearson ha proposto un indice di bontà di adattamento, che si basa proprio sulla differenza fra frequenze empiriche e frequenze teoriche :

$$X^2 = \sum_{i=1}^{k} \frac{\left(n_i - n_i^*\right)^2}{n_i^*}$$

Ovviamente, quanto più il valore di X^2 si avvicina a 0, tanto migliore sarà l'adattamento:

probabilità teoriche p_i	frequenze assolute teoriche $n_i^* = Np_i$	$n_i - n_i^*$	$(n_i - n_i^*)^2$	$(n_i - n_i^*)^2 / n_i^*$
0,0033	0,2621	0,7379	0,5445	2,0772
0,0219	1,7498	0,2502	0,0626	0,0358
0,0692	5,5331	0,4669	0,2180	0,0394
0,1377	11,0164	-2,0164	4,0660	0,3691
0,1935	15,4826	-1,4826	2,1980	0,1420
0,2040	16,3195	5,6805	32,2685	1,9773
0,1672	13,3790	-1,3790	1,9017	0,1421
0,1091	8,7299	-1,7299	2,9927	0,3428
0,0575	4,6009	0,3991	0,1593	0,0346
0,0247	1,9758	-0,9758	0,9521	0,4819
0,0087	0,6942	0,3058	0,0935	0,1347
≅1				5,7769

Stabiliamo una regola empirica, per cui se X^2 risulta inferiore a (k-1), dove k è il numero dei valori assunti dalla variabile X, allora possiamo ritenere buono l'adattamento. Il valore di X^2 nel nostro caso risulta:

$$X^2 = \sum_{i=1}^{k} \frac{\left(n_i - n_i^*\right)^2}{n_i^*} = 5{,}7769\,,$$

dunque possiamo ritenere che il modello binomiale si adatta bene alla distribuzione empirica osservata.

Consideriamo adesso un altro esempio. In una gara di tiro con l'arco è stata osservata la seguente distribuzione di frequenze del *numero dei centri colpiti* dai 200 arcieri partecipanti:

n. di centri colpiti x_i	frequenze assolute osservate n_i
0	4
1	7
2	17
3	28
4	37
5	33
6	29
7	18
8	14
9	8
≥10	5
totali	200

Osserviamo anche in questo caso una variabile statistica discreta, ma questa volta la nostra scelta verte sul modello di Poisson, per i seguenti motivi:

- la variabile osservata descrive un evento raro, in quanto colpire il centro di un bersaglio non è semplice, come si evince anche dalla basse frequenze associate ai valori più alti della variabile;
- la variabile osservata assume un'infinità numerabile di valori;
- la media e la varianza empiriche sono molto vicine tra loro:

n. di centri colpiti x_i	frequenze assolute osservate n_i	$x_i n_i$	x_i^2	$x_i^2 n_i$
0	4	0	0	0
1	7	7	1	7
2	17	34	4	68
3	28	84	9	252
4	37	148	16	592
5	33	165	25	825
6	29	174	36	1044
7	18	126	49	882
8	14	112	64	896
9	8	72	81	648
≥10	5	50	100	500
totali	200	972		5714

$$M = \frac{\sum_{i=1}^{n} x_i n_i}{N} = \frac{972}{200} = 4{,}86 \qquad \sigma^2 = \frac{\sum_{i=1}^{n} x_i^2 n_i}{N} - M^2 = \frac{5714}{200} - (4{,}86)^2 = 4{,}9504\,.$$

Scelto, quindi, il modello di Poisson, ne stimiamo il parametro λ utilizzando la media e la varianza empirica:

$$\hat{\lambda} = M \cong \sigma^2 \cong 4{,}9\,.$$

Possiamo, dunque, calcolare le probabilità teoriche e l'indice X^2:

$$p_i = p(x) = \frac{\hat{\lambda}^x e^{-\hat{\lambda}}}{x!} = \frac{(4,9)^x e^{-4,9}}{x!}$$

Probabilità teoriche p_i	frequenze assolute teoriche $n_i^* = Np_i$	$n_i - n_i^*$	$(n_i - n_i^*)^2$	$(n_i - n_i^*)^2 / n_i^*$
0,0074	1,4893	2,5107	6,3035	4,2325
0,0365	7,2977	-0,2977	0,0886	0,0121
0,0894	17,8792	-0,8792	0,7731	0,0432
0,1460	29,2028	-1,2028	1,4467	0,0495
0,1789	35,7734	1,2266	1,5046	0,0421
0,1753	35,0579	-2,0579	4,2350	0,1208
0,1432	28,6306	0,3694	0,1364	0,0048
0,1002	20,0414	-2,0414	4,1675	0,2079
0,0614	12,2754	1,7246	2,9743	0,2423
0,0334	6,6833	1,3167	1,7338	0,2594
0,0283	5,6690	-0,6690	0,4476	0,0789
1				5,2937

dove $P(X \geq 10) = 1 - \sum_{x=0}^{9} p(x) = 1 - 0{,}9717 = 0{,}0283$.

Poiché, in questo caso, è

$$X^2 = \sum_{i=1}^{k} \frac{(n_i - n_i^*)^2}{n_i^*} = 5{,}2937,$$

si può ritenere che il modello di Poisson si adatta bene ai dati osservati.

Supponiamo adesso di aver osservato una variabile quantitativa continua:

classi x_i-x_{i+1}	frequenze assolute osservate n_i
<5	8
5-10	10
10-15	23
15-20	30
20-25	18
>25	11
totale	100

L'unico modello teorico che possiamo provare ad adattare, fra quelli visti, è il modello di Gauss, occorre perciò stimarne i parametri μ e σ. A tal fine, possiamo calcolare la media e lo scarto quadratico medio sulla distribuzione empirica:

classi x_i-x_{i+1}	frequenze assolute osservate n_i	valori centrali ${}_cx_i$	${}_cx_in_i$	${}_cx_i^2$	${}_cx_i^2n_i$
<5	8	2,5	20	6,25	50
5-10	10	7,5	75	56,25	562,5
10-15	23	12,5	287,5	156,25	3593,75
15-20	30	17,5	525	306,25	9187,5
20-25	18	22,5	405	506,25	9112,5
>25	11	27,5	302,5	756,25	8318,75
totale	100		1615		30825

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i n_i}{N} = \frac{1615}{100} = 16{,}15$$

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 n_i}{N} - \hat{\mu}^2} = \sqrt{\frac{30825}{100} - (16{,}15)^2} = 6{,}89 .$$

Sappiamo che $P(x_i<X<x_{i+1})=P(z_i<Z<z_{i+1})=P(Z<z_{i+1})-P(Z<z_i)=F(z_{i+1})-F(z_i)$. Occorre, dunque, standardizzare i valori:

$$z_{i+1} = \frac{x_{i+1} - \hat{\mu}}{\hat{\sigma}} = \frac{x_{i+1} - 16{,}15}{6{,}89}$$

e calcolare, servendosi del prontuario delle probabilità integrali di una curva normale standardizzata, la funzione di ripartizione $F(z_{i+1})$ al variare di z_{i+1}:

valori standardizzati z_{i+1}	funzione di ripartizione $F(z_{i+1})$	Probabilità Teoriche $F(z_{i+1})-F(z_i)$	frequenze assolute teoriche n_i^*	$n_i-n_i^*$	$(n_i-n_i^*)^2$	$(n_i-n_i^*)^2/n_i^*$
-1,62	0,0528	0,0528	5,28	2,72	7,3984	1,4012
-0,89	0,1860	0,1332	13,32	-3,32	11,0462	0,8291
-0,17	0,4337	0,2477	24,77	-1,77	3,1275	0,1263
0,56	0,7118	0,2781	27,81	2,19	4,7859	0,1721
1,28	0,9005	0,1887	18,87	-0,87	0,7511	0,0398
∞	1	0,0995	9,95	1,05	1,1049	0,1111
		1				2,6795

Calcolate le probabilità teoriche $F(z_{i+1})-F(z_i)$, le moltiplichiamo per il totale delle osservazioni N=100, per ottenere le frequenze assolute teoriche n_i^*. L'indice X^2 è:

$$X^2 = \sum_{i=1}^{k} \frac{(n_i - n_i^*)^2}{n_i^*} = 2{,}68 ,$$

che risulta minore di (*k-1*), dove, in questo caso, *k=6* è il numero delle classi. Poiché risulta $X^2<5$,si può ritenere che la distribuzione di Gauss descrive bene la distribuzione osservata.