Variabili aleatorie continue

Per descrivere la distribuzione di una variabile aleatoria continua, non si può più assegnare una probabilità positiva ad ogni valore possibile. Si assume allora di poter specificare una funzione, detta **funzione densità di probabilità**, f(x), definita sull'insieme D⊆R di valori possibili, che sia **non negativa** dappertutto e per la quale si abbia

$$\int_D f(x)dx = 1$$

Variabili aleatorie continue: funzione di densità

Tale funzione esprime il modo di distribuirsi della probabilità totale sull'insieme dei valori assumibili da X. Per calcolare la probabilità P(X∈A), per un dato sottoinsieme A di D, occorre procedere per integrazione:

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

Variabili aleatorie continue:funzione di ripartizione

In modo anche più semplice e immediato è possibile definire la **funzione di ripartizione**:

$$F(x)=P(X\leq x)= \int_{-\infty}^{x} f(y)dy$$

Per qualunque intervallo [a,b] si ha, allora:

$$P(a\leq X\leq b)= \int_{a}^{b} f(y)dy = F(a)-F(b)$$

Variabili aleatorie continue: valor medio

Nel caso continuo il valor medio si definisce:

$$E(X)= \int_D xf(x)dx$$

Dove D è il dominio di f

E' immediato verificare, che tale media soddisfa le proprietà della media dimostrate nel caso discreto.

Variabili aleatorie continue: varianza

La varianza, che abbiamo definito come $E[(X - E(X))^2]$, risulta data da:

$$\sigma^2(X) = \int_D (x - E(X))^2 f(x)\,dx$$

e, applicando la proprietà di linearità dell'integrale, si ha ancora:

$$\sigma^2(X) = E(X^2) - E(X)^2.$$

Variabili aleatorie continue standardizzate

Anche per le variabili continue, si verifica che, come nel caso statistico, la variabile Z così definita:

$$Z = \frac{X - E(X)}{\sigma(X)}$$

ha media pari a 0 e varianza e scarto standard pari a 1.

Tale **variabile** si dice **standardizzata**.

Distribuzione uniforme

Sei in attesa di un autobus che passa esattamente ogni 15 minuti. Non conosci i tempi di passaggio dell'autobus. Se l'autobus non passa entro 5 min perderai il treno. Che probabilità hai di farcela?

Costruiamo la funzione di densità di probabilità $f(x)=c$ per $0 \leq x \leq 15$, $f(x)=0$ altrove; poiché $\int_0^{15} c\,dx=1$, si ottiene $c=1/15$. La probabilità che l'autobus passi entro 5 min è dunque $\int_0^5 1/15\,dx=5/15=1/3$

Distribuzione uniforme

In generale, la distribuzione uniforme su un intervallo $[a, b]$ ha come funzione di densità di probabilità

$f(x)=c$ per $a \leq x \leq b$, $f(x)=0$ altrove; poiché $\int_a^b c\,dx=1$, si ottiene $c=1/(b-a)$.

Costruisci la funzione di ripartizione:

$P(X \leq x)=F(x)=0$ se $x<a$, $F(x)=(x-a)/(b-a)$ per $a \leq x \leq b$, infine $F(x)=1$ per $x>b$

Si ha $E(X)= \int_a^b x/(b-a)dx=(a+b)/2$, $V(X)=(b-a)^2/2$

Verificalo per esercizio! Quanto vale la mediana?

Distribuzione esponenziale

Vogliamo studiare il tempo X che intercorre tra due eventi consecutivi in un fenomeno di Poisson.

Indichiamo con a il valor medio di eventi in un intervallo di tempo unitario. Allora in un tempo t avremo in media at eventi. Sia Y(t) la v.a. discreta che conta il numero di eventi nell'intervallo di tempo [0, t], abbiamo $P(Y(t)=k)= at^k e^{-at}/k!$

Dire che il primo evento avviene dopo un tempo t>0 è equivalente a dire che non ci sono eventi nell'intervallo [0, t], $P(X>t)=P(Y(t)=0)=e^{-at}$ $\forall t \geq 0$

Distribuzione esponenziale

D'altra parte il primo evento è sicuramente dopo il tempo 0, per cui $P(X<t)=0$ per $t<0$. Poiché $F_X(t)=P(X\leq t)=1-P(X>t)$; dunque la funzione di ripartizione della distribuzione esponenziale risulta

$F_X(t)=1- e^{-at}$ se $t\geq 0$, $F_X(t)=0$ se $t<0$;

Derivando otteniamo la funzione di densità di probabilità:

$f_X(t)= a e^{-at}$ se $t>0$, $f_X(t)= 0$ se $t\leq 0$

Si ottiene $E(X)= 1/a$, $V(X)=1/a^2$

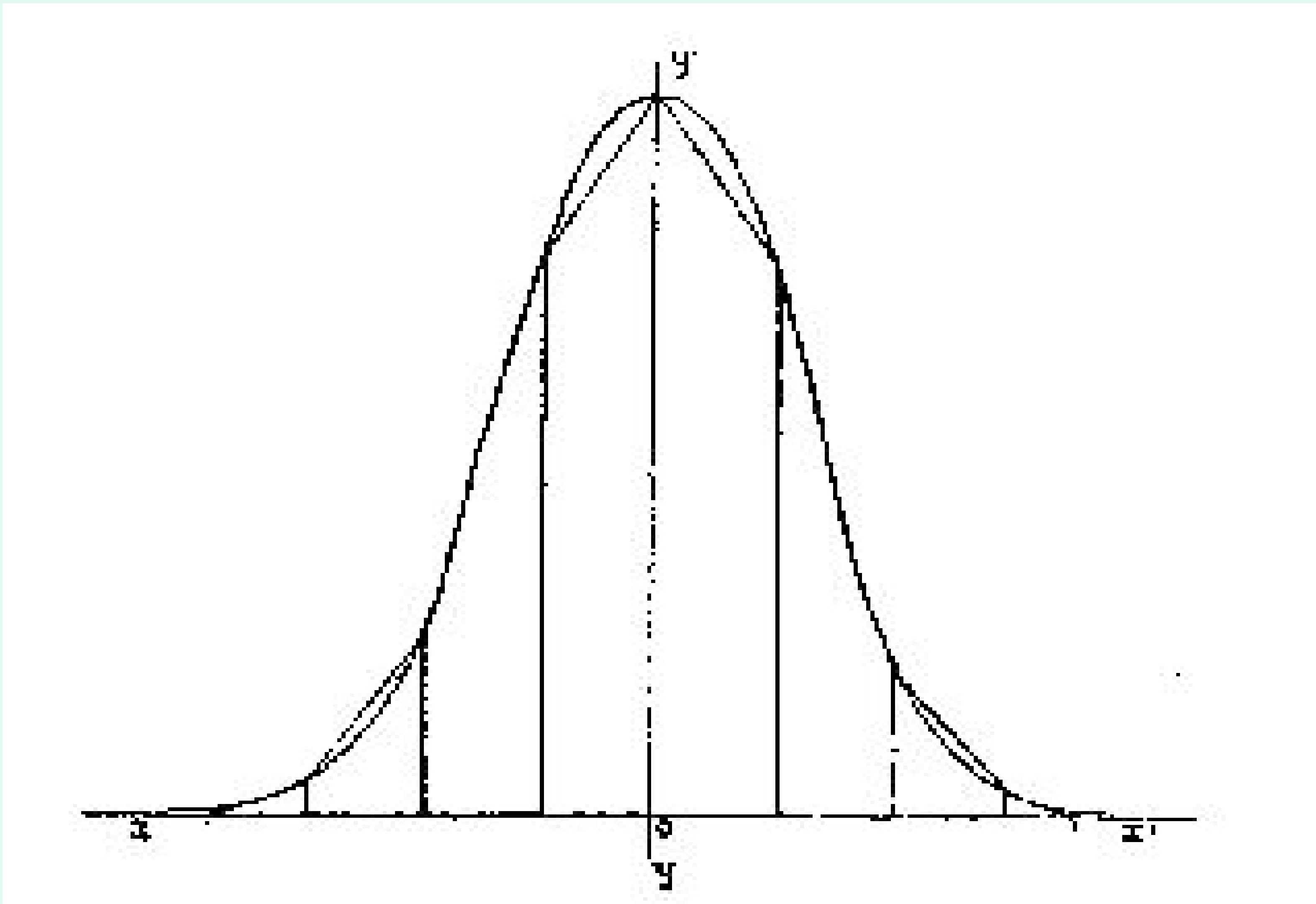
Verificalo per esercizio! Quanto vale la mediana?

Variabili aleatorie gaussiane

La distribuzione normale (riconoscibile dalla curva a forma di campana) è la più usata tra tutte le distribuzioni, perché molte distribuzioni che ricorrono naturalmente sono molto simili ad essa.

La sua derivazione matematica fu presentata per la prima volta da De Moivre nel 1733, ma è spesso riportata come la distribuzione Gaussiana, dal nome di Carl Gauss (1777-1855), che ricavò anche la sua equazione da uno studio degli errori nelle misure ripetute della stessa quantità.

Gaussiana standard



Variabili aleatorie gaussiane

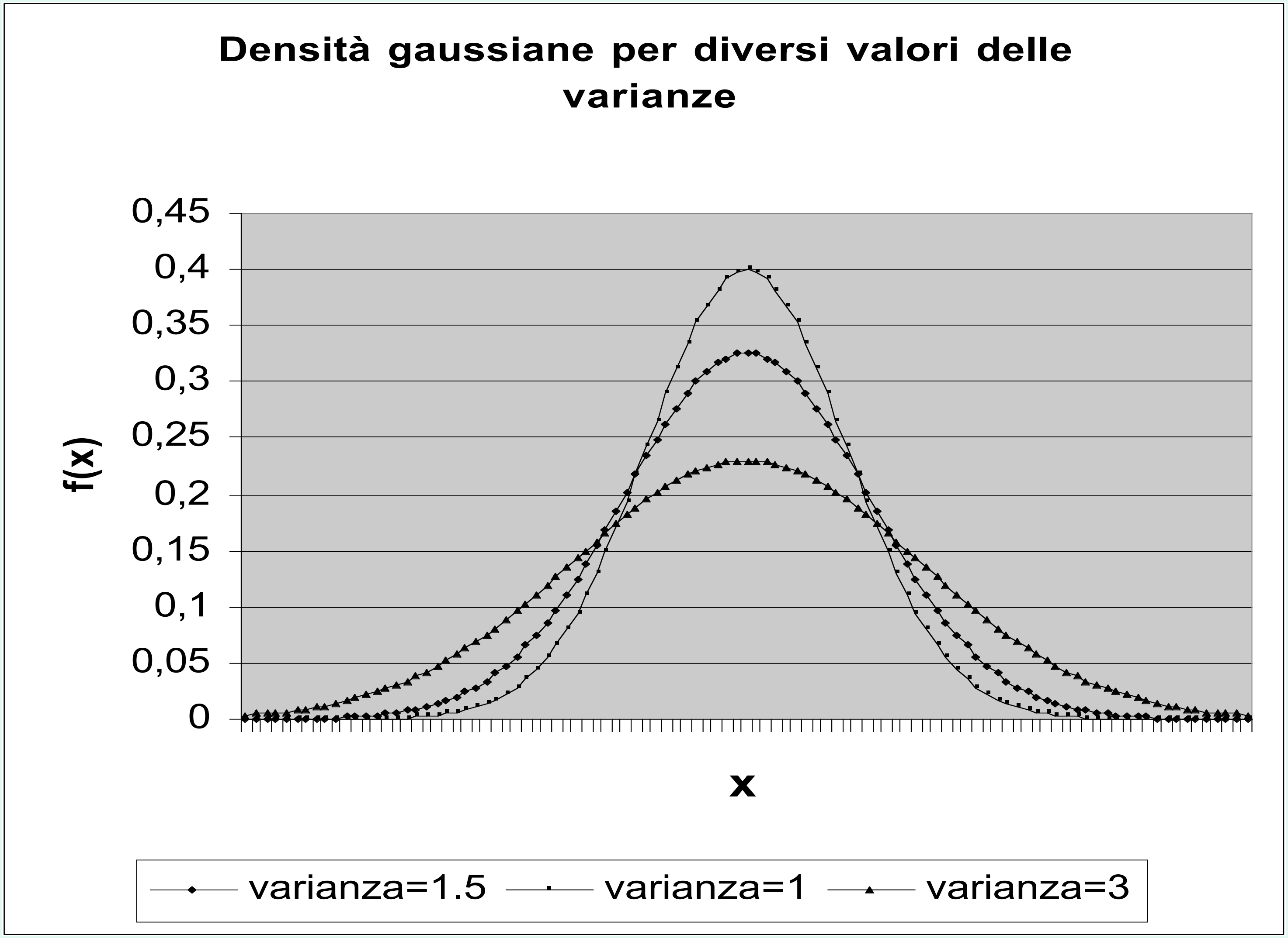
L'espressione della funzione densità della curva normale è:

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

$E(X)=\mu$

$\sigma(X)= \sigma$

Variabili aleatorie gaussiane



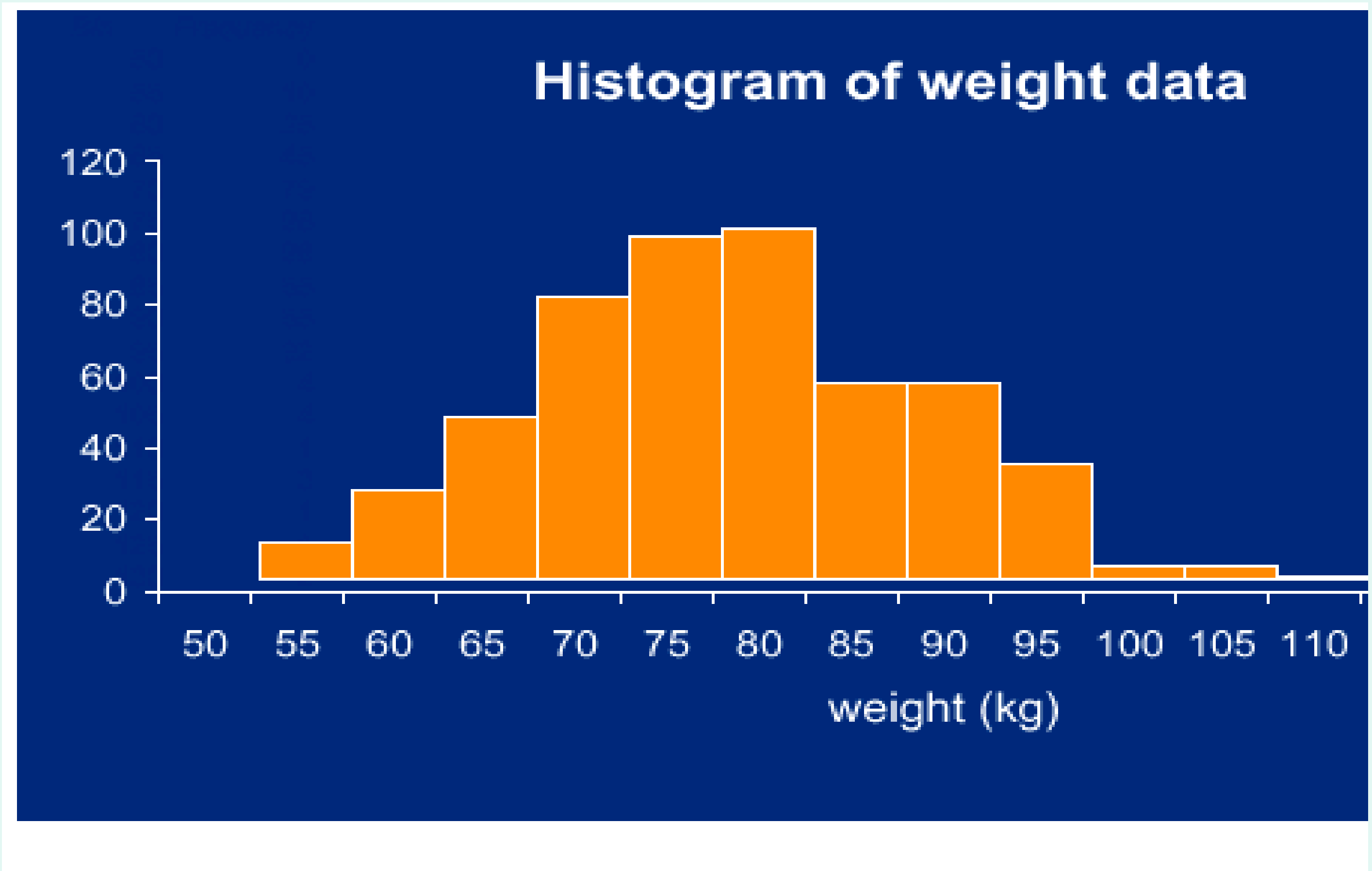
Variabili aleatorie gaussiane

Per una distribuzione normale o quasi normale, eventualmente standardizzando la variabile e utilizzando le tavole della gaussiana standard (media 0 e varianza 1), si osserva che

(a) approssimativamente il 95% di tutti i valori dovrebbe essere compreso entro due deviazioni standard della media.

(b) praticamente tutti i valori dovrebbero essere entro 3 D.S. della media.

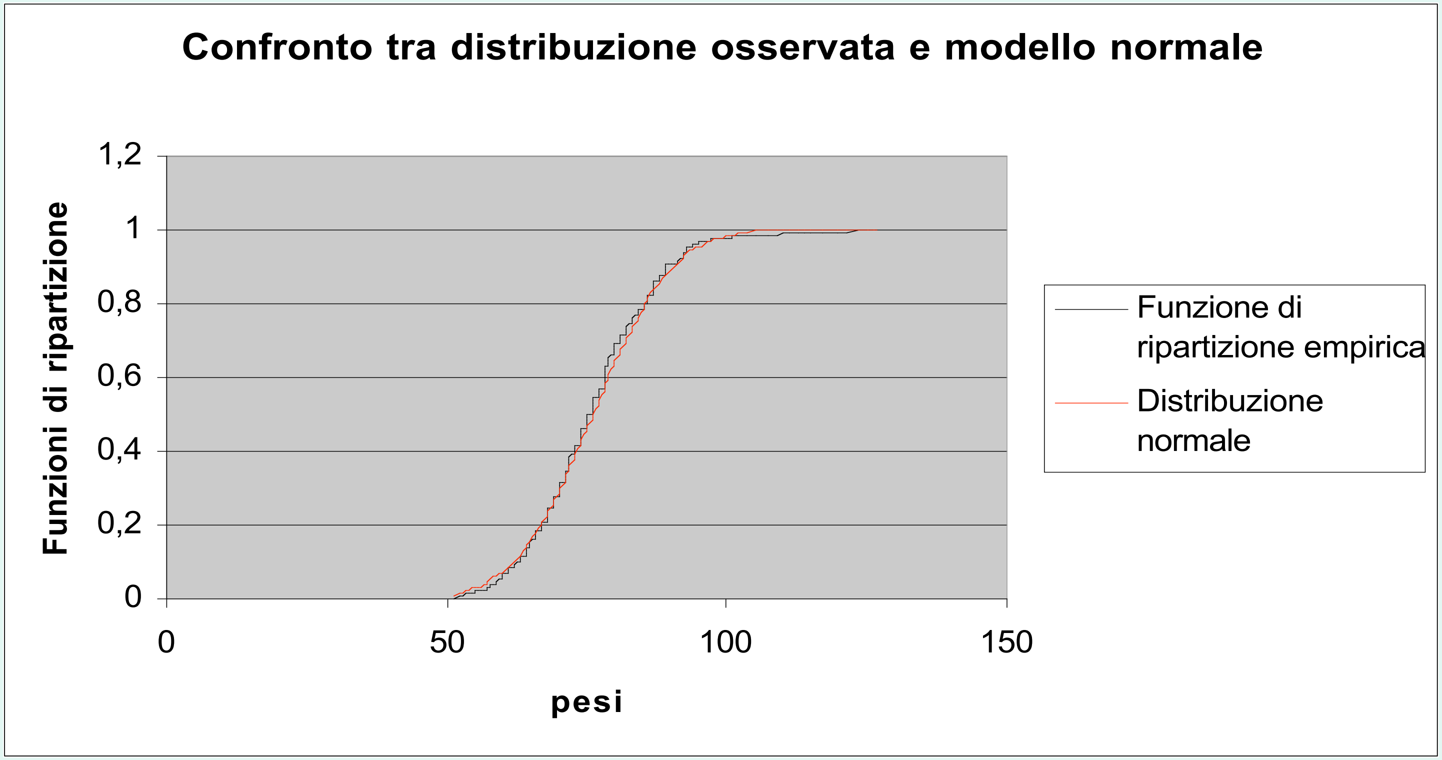
Variabili aleatorie gaussiane e dati empirici



(*materiale didattico Prof.Carla Rossi, Università La Sapienza-Roma*)

Variabili aleatorie gaussiane e dati empirici

Il modello normale con stessa media (76,17) e stessa deviazione standard (11,08) approssima bene la funzione di ripartizione empirica.



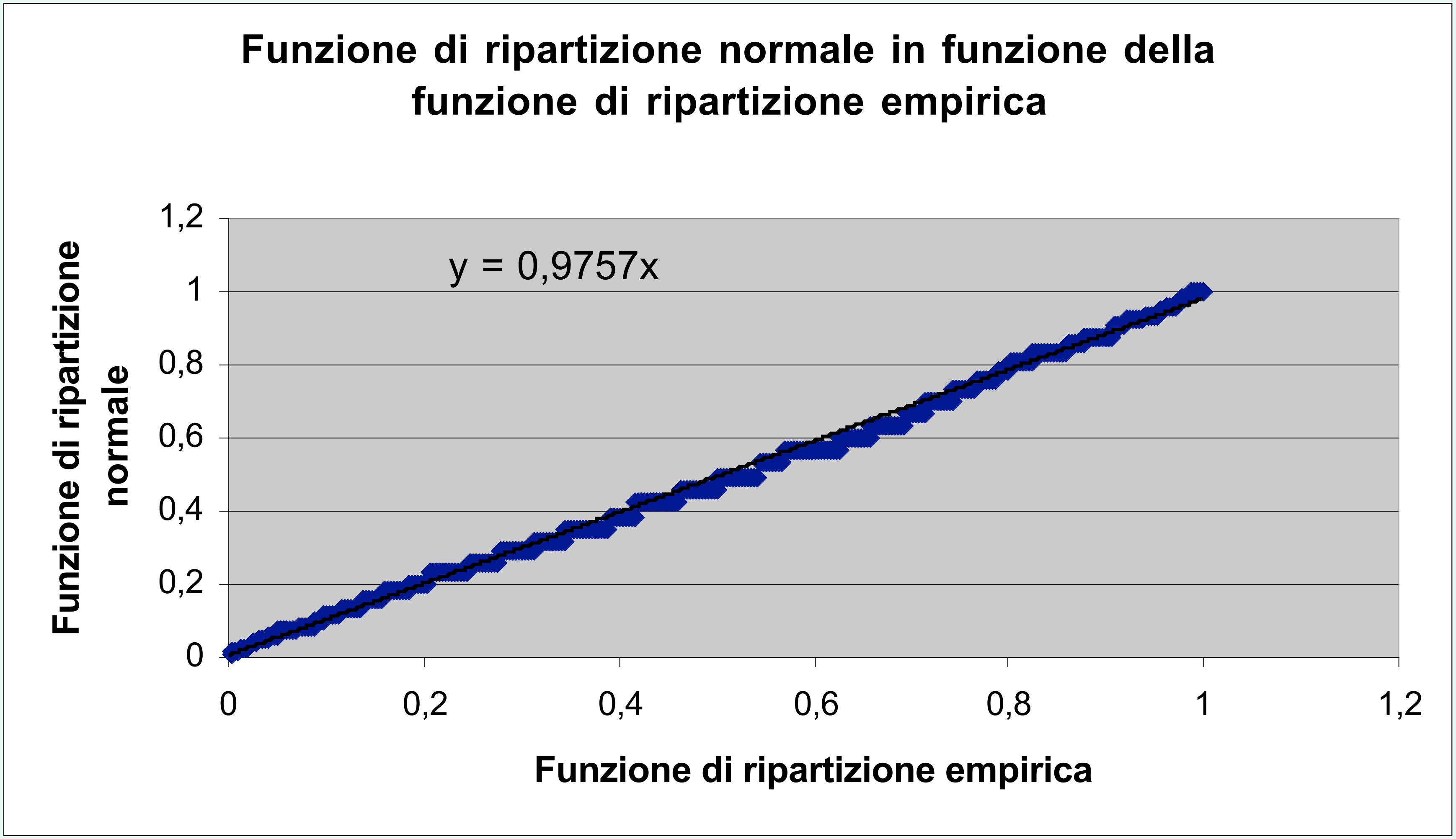
Variabili aleatorie gaussiane e dati empirici

Per meglio confrontare le due funzioni di ripartizione riportiamo i loro corrispondenti valori su un piano cartesiano:

Per ogni valore di x osservato, consideriamo la funzione di ripartizione empirica $F^*(x)$ e la funzione di ripartizione teorica (normale) $F(x)$ e rappresentiamo nel piano il punto che ha per ascissa $F^*(x)$ e per ordinata $F(x)$.

Se il modello approssima bene la distribuzione empirica i punti si addensano attorno alla diagonale del primo quadrante e sono bene interpolati dalla bisettrice con equazione $y=x$.

Variabili aleatorie gaussiane e dati empirici: P-plot



Teorema del limite centrale

Il teorema afferma che, se si ha un certo numero di **variabili aleatorie indipendenti X_i (i=1,2,…,n) con la stessa media m e la stessa varianza σ^2**, allora la successione di variabili aleatorie:

$$Z_n = (Y_n - m)/\sqrt{(\sigma^2/n)}$$

con:

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

tende in distribuzione ad una variabile aleatoria con distribuzione gaussiana standardizzata.

Teorema del limite centrale

La v.a.

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

ha valor medio μ e varianza σ^2/n .

La media di n dati campionari indipendenti $(y_1+\ldots+y_n)/n$ verrà interpretata come se fosse la v.a. Y_n, per n grande; essa non avrà esattamente valore μ, ma ci aspettiamo che non sia troppo lontana da questo valore, essendo la varianza $\sigma^2/n \to 0$ per $n \to \infty$