Dal discreto al continuo

Le distribuzioni di probabilità già viste servono a rappresentare delle caratteristiche discrete

L'insieme dei valori possibili è formato da punti isolati che possono essere "contati" cioè posti in corrispondenza biunivoca con l'insieme dei numeri naturali.

Sono inadatte per descrivere il comportamento di quantità i cui valori ricadono in un intervallo di numeri reali: Distanze,pesi, altezze, etc.

Per affrontare questi aspetti è necessario ampliare il nostro bagaglio di strumenti d'analisi

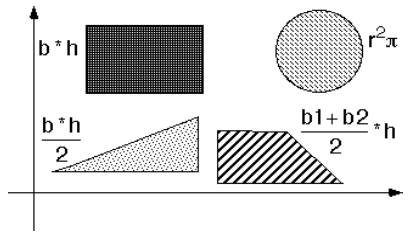
Variabili casuali continue

Le V.C. continue assumono valori in un intervallo di numeri reali (a,b).

Non è possibile osservare PUNTUALMENTE il risultato di una prova, ma solo per intervalli.

La definizione della distribuzione di probabilità delle V.C. continue è DIVERSA da quella delle V.C. discrete. Qui si parla di FUNZIONE DI DENSITA'

Nelle prossime lezioni si farà uso del concetto di area e delle tecniche di integrazione (a livello elementare, ma rigoroso)

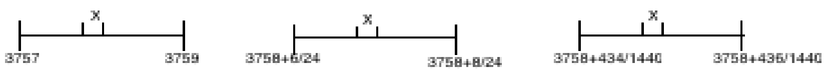


Universo degli eventi continuo

Ciò che una esperienza può provare è che il valore di una variabile continua sia il valore centrale di un intervallo di ampiezza determinata dalla precisione degli strumenti:

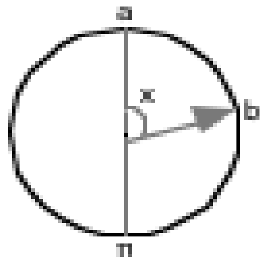
$$X_i - \frac{\Delta X_i}{2} \leq X_i \leq X_i + \frac{\Delta X_i}{2}$$

All'aumentare della capacità di dettaglio, gli estremi dell'intervallo sqranno più vicini rimanendo però sempre distinti.



Esempio

Un esperimento consiste nel girare una freccia lungo una ruota in senso orario:



L'evento è l'intervallo E={x|a≤x≤b} formato dal punto finale "b" raggiunto dalla freccia (dopo un'impulso casuale) con il punto iniziale "a"; l'universo degli eventi è S={x|0≤x≤2π}.

Se è vera l'equiprobabilità è ragionevole assegnare:

P(E)=Arco(b-a)/2π

cioè come rapporto tra arco dell'evento a favore ed arco degli eventi possibili

Inconvenienti/1

Assegnare probabilità per intervalli è possibile, ma non ci possiamo basare sulla identità Intervallo=Evento perché crea problemi:

L'intersezione di una successione finita o enumerabile di intervalli è un intervallo, ma l'unione di due intervalli non sempre è un intervallo.

Siano : a=x_0 < x_1< x_2 < … <x_{n-1} < x_n=b punti distinti di R. Allora:

$$\bigcup_{i=1}^{n}\left(x_{i-1},x_i\right]=(a,b] \;\; con \;\; \left(x_{i-1},x_i\right]\cap\left(x_i,x_{i+1}\right]=\varnothing; \;\; \bigcup_{i=1}^{n}\left[x_{i-1},x_i\right]=[a,b] \;\; con \;\; \left[x_{i-1},x_i\right]\cap\left[x_i,x_{i+1}\right]\neq\varnothing$$

$$\bigcup_{i=1}^{n}\left(x_{i-1},x_i\right)\neq(a,b)$$

per ottenere l'uguaglianza occorre aggiungere i punti singoli interni che non sono intervalli.

Inconvenienti/2

Tutte le σ-algebre sono incluse tra i seguenti casi estremi

$$La \;\; più \;\; povera: \;\; W_* = \{\varnothing, S\}; \qquad La \;\; più \;\; ricca: \;\; W^* = \{E | E \supseteq S\}$$

Se di S è finito o enumerabilmente infinito la W^* è gestibile con i postulati di Kolmogorov (estesi agli eventi enumerabili).

Se invece S è continuo, W^* è troppo ampia per poter essere probabilizzata nel suo complesso e W_* è troppo semplice per essere utile.

Esiste una σ-algebra abbastanza grande da contenere gli eventi di un qualche interesse e, nello stesso tempo, a tanto piccola da non creare problemi quando si applichino i postulati di Kolmogorov?

La classe di Borel

Poniamo S=R (asse reale (-∞, +∞)

Quale può essere l'elemento base della σ-algebra per questo universo?

Scegliamo l'intervallo semiaperto a sinistra e chiuso a destra.

$$\left(a_i,b_i\right]=\left\{x\in R | a_i < x \le b_i\right\}; \quad a_1 < a_2 < \ldots < a_n; \;\; b_1 < b_2 < \ldots < b_n \;\; -\infty \le a_1; \;\; b_n < \infty$$

Qui (a,∞] è inteso come (a,∞) e il complemento di (-∞,b] è pure un intervallo semiaperto a sinistra e chiuso a destra.

A partire da questi "mattoni", costruiremo un'algebra con le operazioni di unione, intersezione e negazione di intervalli disgiunti del tipo: (,]

$$Se \;\; \left(a_i,b_i\right],\left(a_j,b_j\right]\in S \;\; con \;\; \left(a_i,b_i\right]\cap\left(a_j,b_j\right]=\varnothing \;\; per \;\; i\neq j \;\; \Rightarrow \;\; E_n=\bigcup_{i=1}^{n}\left(a_i,b_i\right]\in W$$

La classe di Borel/2

Per ottenere una σ-algebra dobbiamo aggiungere i limiti delle loro successioni monotone.

Condizione necessaria e sufficiente affinché un'algebra diventi una σ-algebra è che essa includa anche le successioni crescenti o decrescenti di eventi.

In tal caso W è la σ-algebra B minimale dell'asse reale (classe di Borel)

PROBLEMA

L'evento E={xε[0,1]|x è un irrazionale} non è un intervallo cioè non è un elemento della classe di Borel per cui il sistema assiomatico non è in grado di dare una probabilità al suo verificarsi.

Esempio

Verifichiamo che la seguente espressione descriva una funzione di densità:

$$h(x) = \begin{cases} \dfrac{3x^2 + 2x + 1}{36} & 1 \le x \le 3 \\ 0 & altrove \end{cases}$$

Notiamo subito che la funzione è positiva nel supporto. Controlliamo l'area sottesa:

$$\int_1^3 \frac{3x^2 + 2x + 1}{36} dx = \frac{1}{36}\left[\int_1^3 3x^2 dx + \int_1^3 2x dx + \int_1^3 1 dx\right] =$$

$$= \frac{1}{36}\left[x^3\Big|_1^3 + x^2\Big|_1^3 + x\Big|_1^3\right] = \frac{1}{36}\left[26 + 8 + 2\right] = 1$$

L'integrale si calcola limitatamente al supporto dato che, al di fuori di questo, il contributo della h(.) all'area è nullo.

Spazi di probabilità e Var. Cas. Cont.

Si parte e si arriva nello stesso tipo di dominio.

Dobbiamo però risolvere il problema di assegnare le probabilità agli intervalli secondo le indicazioni di una particolare P(.);

Il calcolo integrale risponde bene a questa esigenza.

Sia S=(0,∞), "B" la classe di Borel di "S" e E1,E2 Ρ B;

$$P(E_1 \in B) = \int_{E_1} e^{-x} dx;\ P(E_1 \in B) = \int_{E_2} e^{-x} dx;\ P(E_1 \cup E_2) = \int_{E_1 \cup E_2} e^{-x} dx = \int_{E_1} e^{-x} dx; + \int_{E_2} e^{-x} dx;$$

E' evidente che P(E)≥0; P(S)=1 e che P(.) è σ-additiva

Inoltre esprime una funzione di insieme come funzione puntuale.

Funzione di insieme e Ripartizione

Esiste corrispondenza biunivoca tra una funzione di insieme σ-additiva: P(.) dello spazio di probabilità ed una funzione di ripartizione

$$1.\ 0 \le F(x) \le 1,\ x \in R; \qquad 2.\ F(-\infty) = 0;\ F(\infty) = 1;$$

$$3.\ F(x) = F(x+) \qquad 4.\ F(x_1) \le F(x_2)\ se\ x_1 < x_2$$

con assegnazione della probabilità in base alla regola

$$P\{a < x \le b,\ x \in R\} = P\{(a,b]\} = P\left\{\bigcup_{i=1}^{n}(a_i, b_i]\right\} \quad a \le a_1 < b_1 \le a_2 < b_2 < \ldots < b_n \le b$$

$$= \sum_{i=1}^{n}\left[F(b_i) - F(a_i)\right] = F(b) - F(a) \quad per \quad (a,b] \in B$$

Considerazioni

Sia S={xεRl0≤ x ≤4} ed E={xεSlx è irrazionale} ad E dobbiamo per forza assegnare probabilità uno dato che nell'intervallo vi sono sicuramente √2 e π. Però l'evento E è diverso dall'evento certo.

La probabilità che la X assuma un valore specifico X=c è:

$$P(X = c) = F(c) - \lim_{x \to c-} F(x) = F(c) - F(x-)$$

la funzione di ripartizione ha un salto se e solo se P(X=c)>0 e che, se "X" è continua allora P(X=c)=0. Inoltre, si ha:

$$P\left[X \in \left(R - \{c\}\right)\right] = 1$$

che ripropone la tematica di eventi con probabilità uno diversi dall'evento certo.

Variabili continue e frequenza

Ciò che una esperienza può provare è che una X continua ha un valore al centro di di un intervallo di ampiezza data dalla precisione degli strumenti:

$$X_i - \frac{\Delta X_i}{2} \leq X_i \leq X_i + \frac{\Delta X_i}{2}$$

Se n →° "n" e se ogni X si presenta solo una volta la frequenza del valore singolo è posta pari a :

$$f(X_i) = 0$$

anche se ciò sarebbe appropriato solo per modalità che non si verificano e invece X_i si è verificata;

Tuttavia, non è distinguibile dallo zero perché scompare davanti all'immenso numero di valori possibili.

Variabili continue e frequenza/2

D'altra parte non è certo che a verificarsi sia stata proprio X_i dato che può essere solo una falsa indicazione strumentale.

Tutte le X nell'intervallo :

$$X_i - \frac{\Delta X_i}{2} \leq X_i \leq X_i + \frac{\Delta X_i}{2}$$

producono $X = X_i$ perché lo strumetno di misurazione non consente la distinzione.

$X = X_i$ non è solo se stessa, ma un intervallo di valori, quindi

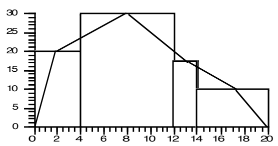
$$f(X_i) > 0$$

piccolo quanto si vuole, ma tale che, sommando le frequenze quasi-zero di ciascuna delle infinite modalità, si ha un numero positivo

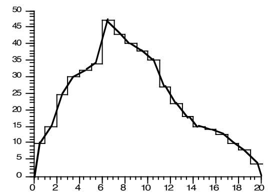
La funzione di densità

ESEMPIO:
Un campione di schede logiche di un elettrodomestico è stato raggruppate in classi annuali di durata:

Durata		Schede	nh_i
0	4	80	50.0
4	12	240	30.0
12	14	35	17.5
14	20	60	10.0
		415	



Durata		Schede	Durata		Schede
0	1	10	10	11	35
1	2	15	11	12	27
2	3	25	12	13	22
3	4	30	13	14	18
4	5	32	14	15	15
5	6	34	15	16	14
6	7	47	16	17	13
7	8	43	17	18	10
8	9	40	18	19	8
9	10	38	19	20	4
					480



Se suddividiamo le classi in sottoclassi di ampiezza unitaria.

L'istogramma sarà formato da rettangoli di base uno e di altezza pari alla frequenza relativa.

inoltre, l'area sottesa al poligono coinciderà con quella totale dell'istogramma.

La funzione di densità/2

Supponiamo ora di estendere la rilevazione all'infinito, ma con classi pari ad un infinitesimo "dx" cioè punti-intervallo:

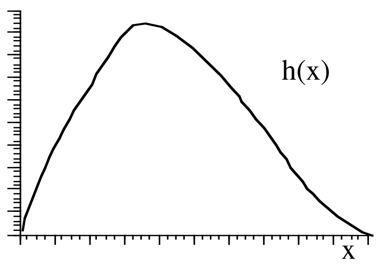
$$[x, \; x + dx]$$

ovvero l'intorno più piccolo diverso dal punto singolo.

L'istogramma non è più tracciabile e diventa simile al poligono di frequenza.

A questo punto è possibile sostituire per interpolazione una curva matematica che esprima per ogni punto-intervallo la densità di frequenza e racchiuda la stessa area.

Tanto più se la curva ricorda una funzione nota



Definizione formale

La funzione di densità di una V.C. continua "X" è una funzione definita su per ogni valore della X

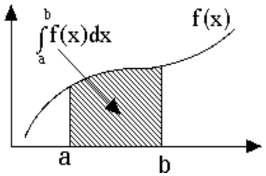
1. $f(X) \geq 0 \quad \text{se} \quad a \leq X \leq b$
2. $f(X) = 0 \quad \text{se} \quad (X < a) \cup (X > b)$
3. $\int_{-\infty}^{\infty} f(X)dx = 1$

Le 1^ deriva dal fatto che la probabilità non può essere negativa.

La 2^ estende a tutto l'asse reale il campo di variazione della V.C. L'intervallo (a,b) è detto SUPPORTO della V.C.

La 3^ deriva dal fatto che l'evento certo deve avere probabilità 1

il simbolo "$\int_a^b$" indica l'integrale della f(X) in (a,b) cioè la misura dell'area sottesa alla curva della funzione f(x) nell'intervallo (a,b)



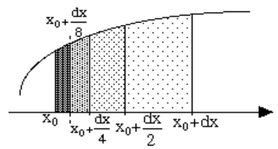
N.B. La f(X) non dà la probabilità di X, ma è proporzionale alla probabilità che X ricada in un intervallo infinitesimo centrato su X.

Conseguenze della definizione

- **La probabilità che la V.C. continua assuma uno specifico valore è zero**

$$P(X = x_0) = P(x_0 \leq X \leq x_0) = \int_{x_0}^{x_0} f(x)dx = 0$$

in quanto l'area di un segmento è zero per definizione



Da notare come il rettangolo costruito intorno a X0 riduca la sua area sempre di più fino ad azzerarsi

- **Nel calcolo di probabilità per intervalli è indifferente includere o no gli estremi**

$$(X \leq x_0) = (X < x_0) \cup (X = x_0) \Rightarrow P(X \leq x_0) = P(X < x_0) + P(X = x_0) = P(X < x_0)$$

questa è zero

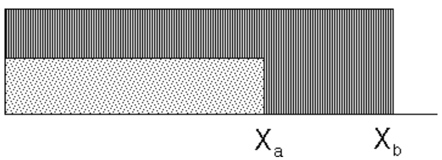
Conseguenze della definizione/2

- **La probabilità che la V.C. assuma valori maggiori di un valore dato si ricava con la probabilità complementare**

$$P[(X > x_0) \cup (X \leq x_0)] = P(X > x_0) + P(X \leq x_0) = 1 \Rightarrow P(X > x_0) = 1 - P(X \leq x_0) = 1 - \int_{-\infty}^{x_0} f(x)dx$$

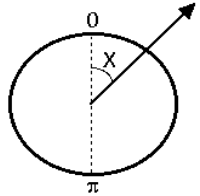
- **Se** $X_a \leq X_b \Rightarrow P(X \leq X_a) \leq P(X \leq X_b)$

Questo discende direttamente dal fatto che: $\{x | x \leq X_a\} \subseteq \{x | x \leq X_b\}$



Esempio

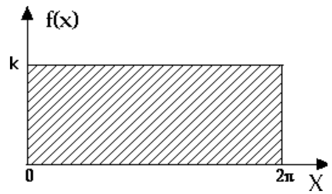
Un esperimento consiste nel girare casualmente una freccia lungo una ruota.



La variabile è l'angolo, in radianti, formato dal punto finale della freccia con il punto iniziale in "0".

Poiché nessun settore del cerchio è privilegiato ogni sottointervallo di (0,2π) ha la stessa probabilità. Questo è possibile solo con una funzione di densità costante (UNIFORME).

$$f(x) = \begin{cases} k & \text{per} \quad 0 \leq x \leq 2\pi \\ 0 & \text{altrove} \end{cases}$$



La scelta di "k" segue dalla condizione che l'area ad essa sottesa è uno:

$$k * 2\pi = 1 \Rightarrow k = \frac{1}{2\pi}$$

Esempio

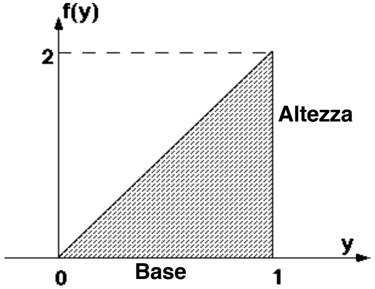
Sia Y una variabile casuale continua con funzione di densità

$$f(y) = \begin{cases} 2y & \text{per } 0 < y < 1 \\ 0 & \text{altrove} \end{cases}$$

Controlliamo che si tratti in effetti di una funzione di densità.

1) Nell'intervallo di definizione è non negativa

2) L'area sottesa alla retta è pari all'area del triangolo con base 1 ed altezza 2. Quindi: Area sottesa=1*2/2 =1



Da notare che, in base alla definizione della funzione di densità, è indifferente includere o escludere gli estremi del supporto. Si può infatti definire in modo equivalente:

$$f(y) = \begin{cases} 2y & \text{per } 0 \leq y \leq 1 \\ 0 & \text{altrove} \end{cases}$$

Altro Esempio

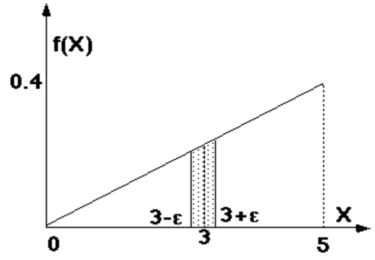
Consideriamo la seguente funzione di densità:

$$f(X) = \begin{cases} 0.08X & \text{per } 0 \leq X \leq 5 \\ 0 & \text{altrove} \end{cases}$$

Calcoliamo la probabilità che X=3

Essa NON è
0.08*3=0.24

Ma è zero



0.24 è la probabilità che la X ricada in un intervallo infinitesimo centrato su X=3.

Caratteristiche delle v.c. continue

La definizione del valore atteso e della varianza delle v.c. continue, ricalca quella delle v.c. discrete:

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx; \qquad \sigma^2 = E\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

Analogamente si possono definire i momenti centrali ed i momenti scarto

$$\mu_r = E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx; \qquad \bar{\mu}_r = E\left[(X-\mu)^r\right] = \int_{-\infty}^{\infty} (x-\mu)^r f(x) dx$$

E' evidente che, anche nel caso delle continue si ha:

$$\sigma^2(X) = \bar{\mu}_2(X) = \mu_2(X) - \mu_1^2(X)$$

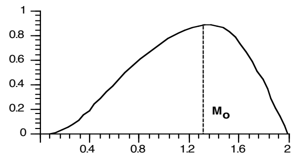
La moda

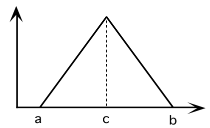
La moda -se esiste- è il valore più probabile ovvero il punto in cui la densità di probabilità raggiunge il suo massimo globale.

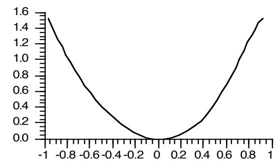
Se la h(x) è dotata di derivata prima per ogni punto nell'intervallo aperto [a,b], la presenza di un massimo in X_0 implicherà che $h'(X_0)=0$.

$$h(x) = \frac{3x^2(2-x)}{4}; \quad 0 < x < 2;$$

$$h'(x) = 3x - \frac{9}{4}x^2 = 0 \Rightarrow x = 0 \ \ o \ \ x = \frac{4}{3}$$







La mediana

La mediana della variabile casuale continua è la soluzione di:

$$0.5 = F(M_e) = \int_{-\infty}^{M_e} h(x)dx \quad ovvero \quad M_e = F^{-1}(0.5)$$

ESEMPIO:

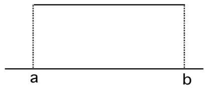
Consideriamo una funzione di densità crescente in [0, 0.5] che esprime la percentuale di astensionismo potenziale in un *referendum*.

$$h(x) = \frac{1}{Ln(2)(1-x)}; \quad 0 < x < 0.5; \quad F(x) = -\frac{Ln(1-x)}{Ln(2)} \Rightarrow M_e = 1 - 0.5^{0.5} = 0.2929$$

Scarto quadratico (densità)

$$\sigma = \sqrt{\int_{-\infty}^{\infty}(x-\mu)^2 h(x)dx} = \sqrt{\int_{-\infty}^{\infty} x^2 h(x)dx - \mu^2}$$

Esempio:
densità uniforme: $h(x) = \frac{1}{b-a}; \quad a \le x \le b$



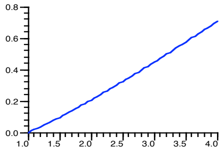
$$\int_a^b \frac{x^2}{(b-a)}dx = \left.\frac{x^3}{3(b-a)}\right|_a^b = \frac{b^3-a^3}{3(b-a)}; \quad \int_a^b \frac{x}{(b-a)}dx = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{b^2-a^2}{2(b-a)}$$

$$= \frac{(b+a)(b-a)}{2(b-a)} = \frac{b+a}{2}$$

$$\sigma^2 = \frac{b^3-a^3}{3(b-a)} - \left[\frac{b+a}{2}\right]^2 = \frac{(b-a)^3}{12(b-a)} - \frac{(b-a)^2}{12} \Rightarrow \sigma = \frac{b-a}{\sqrt{12}}$$

Momenti e densità

Un fenomeno ha densità di frequenza pari allo scarto tra livello raggiunto e la sua radice quadra



$$h(x) = \frac{6(x-\sqrt{x})}{17}; \quad 1 \le x \le 4$$

$$\mu_1 = \frac{6}{17}\int_1^4 x(x-\sqrt{x})dx = \frac{6}{17}\int_1^4 x^2 - x^{3/2}dx = \frac{6}{17}\left[\int_1^4 x^2 dx - \int_1^4 x^{3/2}dx\right] = \frac{6}{17}\left[\left.\frac{x^3}{3}\right|_1^4 - \left.\frac{x^{5/2}}{5/2}\right|_1^4\right]$$

$$= \frac{6}{17}\left[\left(\frac{64}{3}-\frac{1}{3}\right) - \left(\frac{32}{5/2} - \frac{1}{5/2}\right)\right] = \frac{6}{17}\left[\frac{61}{3} - \frac{62}{5}\right] = 2.8$$

$$\mu_2 = \frac{6}{17}\int_1^4 x^2(x-\sqrt{x})dx = \frac{6}{17}\int_1^4 x^3 - x^{5/2}dx = \frac{6}{17}\left[\left.\frac{x^4}{4}\right|_1^4 - \left.\frac{x^{7/2}}{7/2}\right|_1^4\right]$$

$$= \frac{6}{17}\left[\left(64 - \frac{1}{4}\right) - \left(128 - \frac{1}{7/2}\right)\right] = \frac{6}{17}\left[\frac{61}{3} - \frac{62}{5}\right] = 9.693 \qquad \sigma = \sqrt{9.693 - 2.8^2} = 1.36$$

Esistenza (finita) dei momenti

il momento r-esismo esiste finito se e solo se

$$\lim_{x\to\infty} x^r[1 - F(x)] \to 0; \quad \lim_{x\to\infty} x^r[F(-x)] \to 0;$$

maggiore è l'esponente "r" più rapido sarà il decadimento.

Se, per un certo "r", i momenti sono infiniti allora sono infiniti tutti i momenti di ordine superiore perché:

$$\upsilon_r^{s-t} \le \upsilon_t^{s-r}\upsilon_s^{r-t} \Rightarrow \upsilon_{r+1} \le \frac{\upsilon_r^2}{\upsilon_{r-1}}$$

se per una variabile la media risulta infinita sarà infinita anche la varianza e gli altri momenti a seguire.

Se la varianza esiste finita allora anche la media esiste finita.

La asimmetria

Per misurare l'asimmetria delle funzioni di densità si possono usare gli indici già adoperati: l'α_1 e α_2 .

$$h(x)=\begin{cases}\dfrac{2x^2-x}{34.5}; & per \quad 1\le x\le 4\\ 0 & altrove\end{cases}$$

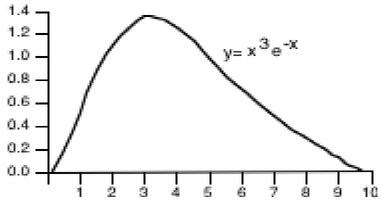
$$M_e = 3.2395,\ \mu = 3.0869,\ \sigma = 0.69547,\ \alpha_1 = -0.2194$$

$$h(x) = x^n e^{-bx} \quad x>0;\quad \mu = \int_0^\infty x^{n+1}e^{-bx}dx = \frac{(n+1)!}{b^{n+2}};$$

$$h'(x) = x^{n-1}e^{-bx}(n-bx) \Rightarrow M_o = \frac{n}{b}$$

$$\sigma = \int_0^\infty x^{n+2}e^{-bx}dx - \mu^2 = \frac{(n+2)!}{b^{n+3}} - \left[\frac{(n+1)!}{b^{n+2}}\right]^2;$$

$$\alpha_2 = \frac{\dfrac{(n+1)!}{b^{n+2}} - \dfrac{n}{b}}{\sigma};\quad nullo \quad se \quad b = \left[\frac{(n+1)!}{n}\right]^{\frac{1}{n+1}}$$



La asimmetria/2

Feller (1950, pp. 78-81) propone il modello arcoseno per situazioni in cui si verificano frequenze elevate nei due estremi della distribuzione e basse al centro quali per esempio le durate dei soggiorni turistici.

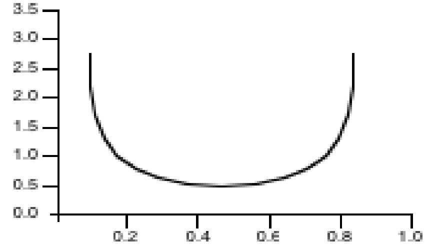
$$h(x) = \frac{1}{\pi\sqrt{x(1-x)}} \quad 0<x<1;\quad F(x) = \left(\frac{2}{\pi}\right) arcoseno\left(\sqrt{x}\right)$$

$$X_p = sen^2\left(\frac{\pi p}{2}\right) \Rightarrow M_e = sen^2\left(\frac{\pi}{4}\right) = \frac{1}{2}$$

$$\mu = \int_0^1 sen^2\left(\frac{\pi p}{2}\right)dp = \frac{2}{\pi}\left\{\frac{x}{4} - \frac{\text{sen}(2x)}{4}\Bigg|_0^{\frac{\pi}{2}}\right\} = \frac{1}{2};\quad S_{M_e} = \frac{1}{\pi};\quad \alpha_1 = \pi\left(\frac{1}{2}-\frac{1}{2}\right) = 0$$

Le misure di asimmetria hanno le stesse finalità e le stesse carenze incontrate nel 3° capitolo.

L'a_1 è nullo se la funzione di densità è simmetrica:



Esempio

Sia X una variabile casuale con funzione di densità:

$$f(X)=\begin{cases}\dfrac{2X^2-X}{34.5} & \text{per} \quad 1<X<4\\ 0 & \text{altrove}\end{cases}$$

Il valore atteso è

$$\mu = \int_1^4 x\left(\frac{2x^2-x}{34.5}\right)dx = \frac{1}{34.5}\int_1^4\left(2x^3-x^2\right)dx = \frac{1}{34.5}\left(\frac{x^4}{2}-\frac{x^3}{3}\right)\Bigg|_1^4 = 3.087$$

Per il calcolo della varianza determiniamo il momento 2°

$$\sigma^2 = E\left(X^2\right) - \mu^2$$

$$E\left(X^2\right) = \int_1^4 x^2\left(\frac{2x^2-x}{34.5}\right)dx = \frac{1}{34.5}\int_1^4\left(2x^4-x^3\right)dx = \frac{1}{34.5}\left(\frac{2x^5}{5}-\frac{x^4}{4}\right)\Bigg|_1^4 = 10.013$$

$$\sigma^2 = 10.013 - 3.087^2 = 0.4834$$

Modelli di variabili casuali

Le probabilità (come tabella o come formula) derivano:

- **Dalle condizioni sperimentali**
- **Dall'aspetto descritto dalla variabile casuale**
- **Da informazioni aggiuntive**

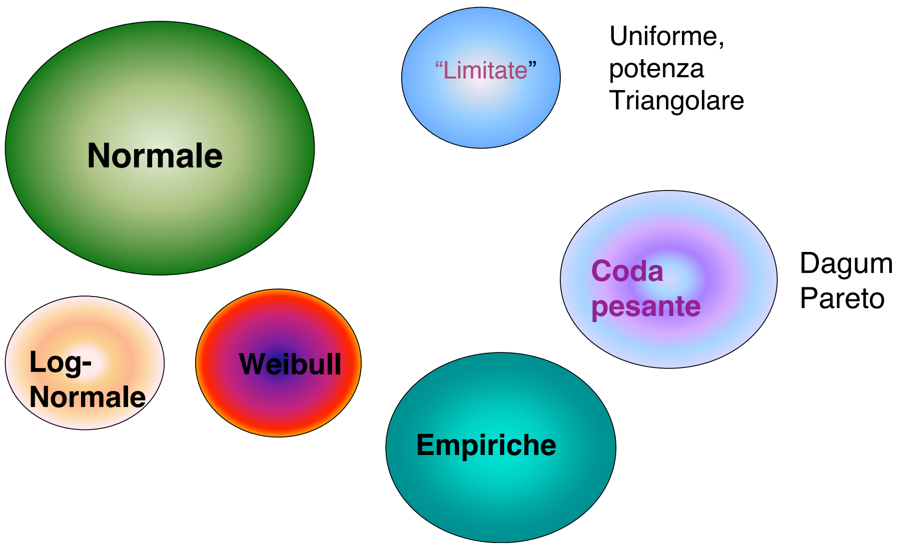
La statistica ha elaborato molti modelli sia per le variabili nominali e discrete che per le variabili continue.

Molti possono essere studiati e approfonditi nel libro di testo (apprendimento libero)

Nel prosieguo analizzeremo un solo tipo di esperimento: estrazione di un campione da una popolazione di unità.

Attenzione! Le variabili casuali sono delle popolazioni.

Breve lista di modelli probabilistici



Nota sulla distribuzione di probabilità

In genere, la distribuzione di un dato aspetto delle unità inserite nella popolazione non può essere ricostruita a partire dalle dalla situazione sperimentale.

Talvolta lo schema probabilistico fornisce distribuzioni del tutto intrattabili.

Compito della Statistica è di definire dei "modelli" di distribuzione e di aiutare a scegliere quello più adatto alla particolare situazione di studio.

Tra i tanti modelli proposti e sviluppati dalla statistica ne studieremo in dettaglio solo pochi

La v.c. Uniforme (Rettamgolare)

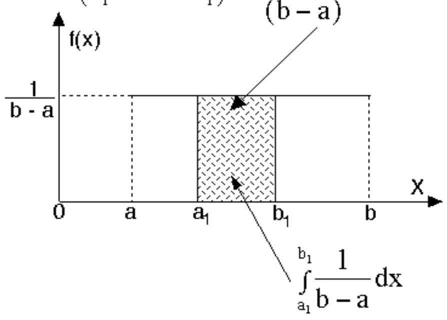
E' la v.c. più semplice e forma la base per lo studio di molte altre v.c.

Le caratteristica principale è che le sue realizzazioni sono equiprobabili

Si applica nelle situazioni in cui il fenomeno:

- **Assume valori in un intervallo limitato [a,b]**
- **La probabilità di ogni sottointervallo di [a, b] è proporzionale all'ampiezza del sottointervallo**

$$P(a_1 \le X \le b_1) = \frac{(b_1 - a_1)}{(b - a)}$$



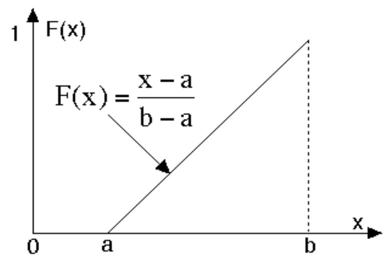
La funzione di densità è:

$$f(X) = \begin{cases} \dfrac{1}{b-a} & \text{per} \quad a \le X \le b \\ 0 & \text{altrove} \end{cases}$$

Uniforme/2

La funzione di ripartizione è pure semplice:

$$F(x) = P(a \le X \le x) = \int_a^x \frac{1}{b-a} dt = \frac{(x-a)}{(b-a)}$$



$$E(X) = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} * \frac{x^2}{2}\Big|_a^b$$

$$= \frac{1}{b-a} * \left(\frac{b^2}{2} - \frac{a^2}{2}\right) = \frac{1}{b-a} * \left[\frac{(b-a)*(b+a)}{2}\right] = \frac{b+a}{2}$$ **= valore atteso della distribuzione**

Nella v.c. UNIFORME il valore atteso coincide con il valore centrale del suo supporto.

Uniforme/3

$$E(X^2) = \int_a^b \frac{x^2}{b-a}dx = \frac{1}{b-a} * \left[\frac{b^3-a^3}{3}\right]$$

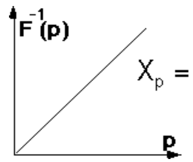
Quantili: $X_p = a + p*(b-a)$

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{(b-a)^2}{12}$$

Esempio:
Supponiamo che la v.c. X abbia la funzione di densità $f(x) = 1 \quad \text{per} \quad 0 < x < 1$

Allora:

$$E(X) = \frac{0+1}{2} = 0.5; \quad \sigma^2(X) = \frac{(1-0)^2}{12} = 0.083$$



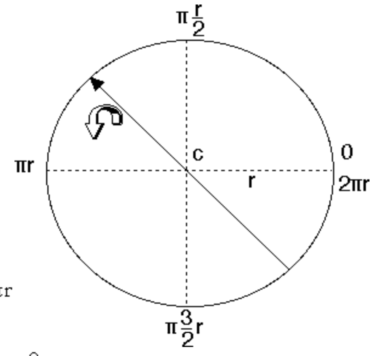
Esempio

Su di un disco di raggio "r" e di centro "c"si fa ruotare una freccia con un impulso causale

Se la freccia è ben equilibrata la posizione"X" in cui si troverà ricalca le condizioni della v.c. uniforme

a) Campo di variazione limitato: [0,2πr]

b) Probabilità proporzionale alla lunghezza dell'arco misurata a partire dallo zero



La funzione di densità di "X" è perciò data da

$$f(X) = \begin{cases} \frac{1}{2\pi r} & \text{per} \quad 0 \le X \le 2\pi r \\ 0 & \text{altrove} \end{cases}$$

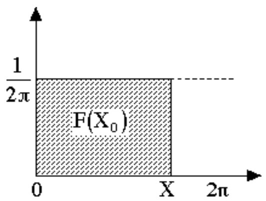
$$E(X) = \frac{2\pi r}{2} = \pi r$$

$$\sigma^2(x) = \left(\frac{\pi}{3}\right)\pi r^2$$

Continuazione esempio

Supponendo il raggio pari ad uno la funzione di densità risulta :

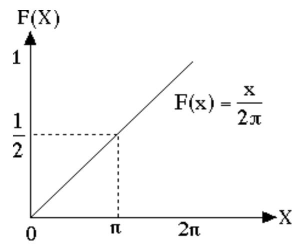
$$f(x) = \begin{cases} \frac{1}{2\pi} & \text{per} \quad 0 \le X \le 2\pi \\ 0 & \text{altrove} \end{cases}$$



La funzione di ripartizione è: $F(x) = \int_{-\infty}^{x} \frac{1}{2\pi} dt = \frac{x}{2\pi}$

Quindi, la probabilità che l'angolo sia inferiore a 180° (cioè inferiore a π) è

$$P(X < \pi) = F(\pi) = \frac{\pi}{2\pi} = \frac{1}{2}$$



Esempio

Gli autobus dell'università passano ogni ora fra le 8.30 e le 13.30.

1) Calcolare la probabilità che una persona, capitando a caso durante tale periodo, debba aspettare almeno un quarto d'ora;
2) Calcolare il tempo medio di attesa ed il suo SQM

La v.c. "X" è "tempo mancante al prossimo bus". Il modello è uniforme dato che "capitare a caso" significa che può capitare in uno qualsiasi dei 60 minuti. Quindi:

$$P(x \ge 15) = 1 - P(X < 15) = 1 - F(15)$$
$$= 1 - \frac{15-0}{60-0} = 1 - 0.25 = 0.75$$

$$E(X) = \frac{0+60}{2} = 30 \ \text{(mezzora)}$$
$$\sigma(X) = \sqrt{\frac{(60-0)^2}{12}} = 17.32 \ \text{minuti}$$

La variabile casuale esponenziale

Questa v.c. si origina dalla Poisson ed è in risposta alla domanda:

Se una sequenza di eventi si verifica nel tempo secondo il modello di Poisson alla media di λ eventi per unità di tempo, quanto tempo bisogna aspettare perchè si verifichi il primo?

Riconsideriamo la Poisson relativamente al problema di arrivi di clienti ad uno sportello bancario

$$f(X) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \quad \textbf{dove} \quad \begin{cases} X & \text{numero di clienti nell' unità di tempo} \\ E(X) = \lambda t & \text{media di arrivi} \\ \sigma^2(X) = \lambda t & \text{varianza degli arrivi} \end{cases}$$

Indichiamo con "T" il tempo trascorso fino all'arrivo del primo cliente ovvero quello che intercorre tra un cliente ed un altro.

"T" è una v.c. continua che assume valori tra zero e infinito

Esempio: durata di un evento

Sia "X" la V.C. che rappresenta la durata di una conversazione telefonica e si adotti come funzione di densità ESPONENZIALE

$$f(x) = e^{-x}$$

Grafico della funzione di densità

La funzione di ripartizione sarà

$$F(x) = 1 - e^{-x}$$

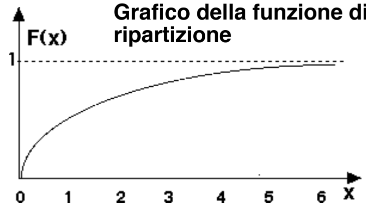


Grafico della funzione di ripartizione

Ad esempio la probabilità che la durata sia tra 2 e 3 è:

$$P(2 \le x \le 3) = F(3) - F(2) = 1 - e^{-3} - \left(1 - e^{-2}\right)$$
$$= e^{-2} - e^{-3} = .1353 - .0498 = 0.0855$$

Esponenziale/2

Analizziamo ora l'evento: il primo arrivo avviene dopo un tempo "t", cioè (T>t)

Tale evento si realizza se non ci sono occorrenze nell'arco di tempo "t", cioè se la v.c. di Poisson è x=0.

Il converso è anche vero: il numero di arrivi durante l'arco di tempo "t" è zero (x=0) se il primo arrivo si verifica a T>t. Quindi

$$P(T > t) = P(X = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$$

Poiché $P(T \le t) + P(T > t) = 1$ **si ha inoltre**

$$P(T \le t) = 1 - e^{-\lambda t}$$

Questa è la funzione di ripartizione della v.c. esponenziale

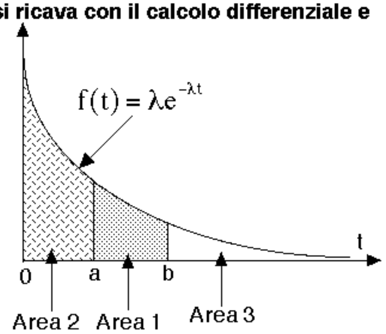
$$F(t) = 1 - e^{-\lambda t} \quad \text{per} \quad t \ge 0$$

Esponenziale/3

La funzione di densità della esponenziale si ricava con il calcolo differenziale e risulta essere

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{per} \quad t \ge 0 \\ 0 & \text{altrove} \end{cases}$$

$$F(t) = 1 - e^{-\lambda t}$$



$$\text{Area } 1 = P(a < T < b) = \int_a^b \lambda e^{\lambda t} dt = F(b) - F(a) = \left(1 - e^{-\lambda b}\right) - \left(1 - e^{-\lambda a}\right) = e^{-\lambda a} - e^{\lambda b}$$

$$\text{Area } 2 = P(T \le a) = F(a) = 1 - e^{-\lambda a} \qquad \text{Area } 3 = P(T > b) = 1 - F(b) = e^{-\lambda b}$$

Esponenziale/4

La funzione di graduazione dell'esponenziale è:

$$X_p = \frac{1}{\lambda} * Ln\left(\frac{1}{1-p}\right)$$

Si dimostra inoltre che

$$E(T) = \frac{1}{\lambda}; \qquad \sigma^2(T) = \frac{1}{\lambda^2}$$

nella esponenziale coincidono media e scarto quadratico medio

Esempio

I clienti di un ipermercato arrivano alle entrate secondo un modello di Poisson ad una media di λ=4 al minuto e cioè

$$P(X = x) = \frac{4^x e^{-4}}{x!}; \quad x = 0,1,2,\ldots,$$

X=numero di clienti in un dato minuto

Quanto tempo occorre aspettare dopo l'apertura prima che entri il primo cliente?

Quanto si deve aspettare tra un cliente e l'altro?

$$P(X = x) = 4e^{-4x} \quad X > 0$$

Ad esempio la probabilità di aspettare meno di mezzo minuto è

$$F(0.5) = 1 - e^{-4*\frac{1}{2}} = 1 - e^{-2} = 86.4\%$$

Esercizio

Si supponga che un rappresentante effettui le consegne secondo una legge di Poisson al ritmo di 4 consegne ogni ora. Cioè

$$P(\text{Rappresentante arriva in t ore}) = \frac{(4t)^x e^{-4t}}{x!} \quad x = 0,1,2,\ldots$$

con una media di λt=4t consegne ogni t ore.

Calcolare la probabilità che un punto vendita aspetti meno di un quarto d'ora prima di vederlo arrivare

Poiché si effettuano 4 consegne all'ora si ha $\lambda = \frac{4}{60} = \frac{1}{15}$ **per cui**

$$P(\text{arriva in meno di t'}) = F(t) = 1 - e^{\frac{t}{15}}$$

Nel caso particolare:

$$P(t \leq 15) = F(15) = 1 - e^{\frac{15}{15}} = 1 - e^{-1} = 0.632$$

Altro esercizio

il monitor di una linea di televisori ha durata descritta da una v.c. esponenziale con media 10 anni.

a) calcolare la probabilità che si guasti dopo 15 anni;
b) si guasti prima del periodo di garanzia di 2 anni.

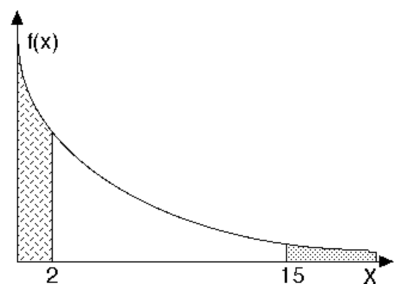
indichiamo con "X" la durata e ricaviamo il parametro "λ"

$$E(X) = 10 \Rightarrow \lambda = \frac{1}{E(X)} = \frac{1}{10} = 0.1$$

a) $P(X > 15) = e^{-0.1*15} = 0.22$

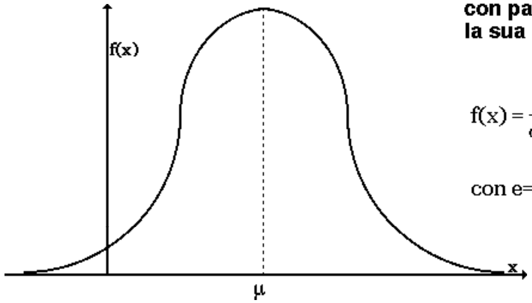
b) $P(X \leq 2) = 1 - e^{-0.1*2} = 0.18$

Il 22% avrà lunga durata, ma il 18% darà dei costi di riparazione in garanzia



La v.c. normale (o gaussiana)

E' la v.c. più importante, più nota e più usata in statistica



Una variabile casuale è di tipo normale con parametri μ e σ : $X \sim N(\mu, \sigma)$ se la sua funzione di densità è

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left\{\frac{-1}{2\sigma 2}[x-\mu]^2\right\}} \quad -\infty<\mu\infty;\ \sigma>0$$

con $e=2.71828;\ \pi = 3.14159$

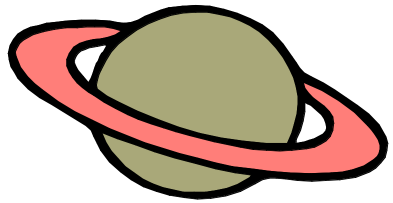
L'importanza di questa v.c. risiede negli indubbi vantaggi formali, ma anche nel fatto che moltissimi fenomeni empirici possono essere rappresentati con un modello di tipo gaussiano

Importanza della normale

Risiede nel fatto che moltissimi fenomeni possono esservi rappresentati.

Infatti, la normale serve da efficace approssimazione di molte altre variabili casuali continue e discrete.

Già nel 17° secolo, Galileo discusse il comportamento delle misurazioni delle distanze astronomiche avendo in mente il modello normale della compensazione tra errori di segno opposto.

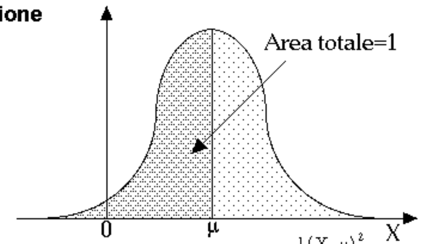


Caratteristiche della Normale

- **La funzione utilizzata è in effetti una funzione di densità**

$$f(X) \geq 0; \qquad \int_{-\infty}^{+\infty} f(X)dX = 1$$

- **La densità diminuisce man mano che si allontana dal centro (L'asse delle ascisse è un asintoto della f(x))**



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

- **la funzione di densità è simmetrica intorno a μ**

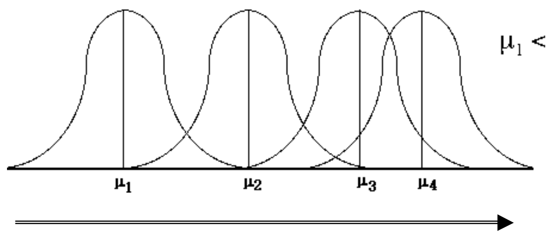
$$f(\mu + x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{\mu+x-\mu}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{\mu-x-\mu}{\sigma}\right)^2} = f(\mu - x)$$

Ne consegue che il valore atteso coincide con la Mediana $Me = X_{50}$

Significato del parametro "μ"

La distribuzione è unimodale e l'ordinata massima si raggiunge per X=μ (La moda)

Quindi, il parametro μ rappresenta il valore più probabilie nonchè il valore atteso e quello che bipartisce il supporto dei valori.
Cambiando μ si modifica la collocazione del grafico



$\mu_1 < \mu_2 < \mu_3 < \mu_4$

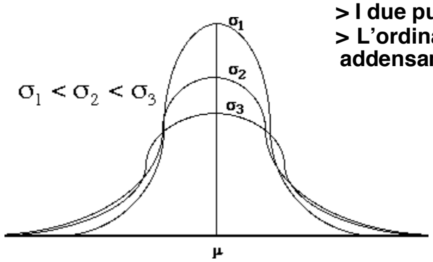
Al variare di μ il grafico resta inalterato nella sua forma. Si modifica solo la sua localizzazione: più a destra se μ aumenta; più a sinistra se μ diminuisce

Significato del parametro "σ"

Il σ corrisponde allo scarto quadratico medio.
La curvatura del grafico della distribuzione normale cambia due volte inflessione in corrispondenza dei punti $x=\mu \pm \sigma$. Inoltre

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

Al diminuire di σ :
> I due punti di flesso tendono ad accentrarsi;
> L'ordinata massima aumenta a causa del maggiore addensamento intorno al centro della distribuzione.



La funzione di ripartizione

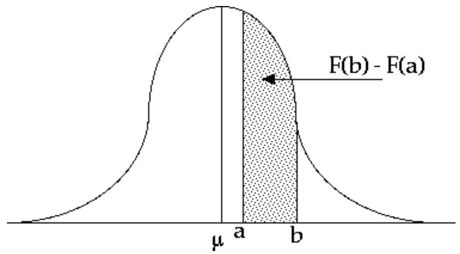
La formula della Normale è complicata e l'integrale

$$F(x) = P(X \le x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

è calcolato con metodi di approssimazione numerica

La v.c. Normale è completamente specificata dai parametri μ e σ cioè noti questi non c'è altro da sapere. Ne consegue che ogni calcolo della F(x) deve essere ripetuto per ogni combinazione dei due parametri

Esiste comunque una tecnica che ci permette ci calcolare le probabilità di tutte le Normali usando una sola tabella

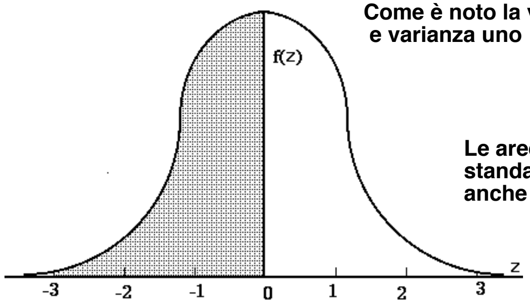


Curva normale standardizzata

E' possibile esprimere la v.c. Normale in unità standard $z = \frac{x - \mu}{\sigma}$

La funzione di densità è ora $f(z) = \frac{1}{\sqrt{2\pi}} e^{\left[\frac{-z^2}{2}\right]}$ **Indicata con N~(0,1)**

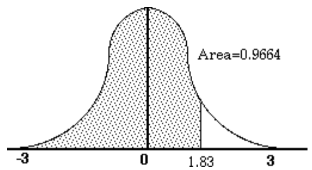
Come è noto la v.c. standardizzata ha media zero e varianza uno



Le aree sottese alla curva normale standardizzate sono state tabulate (sono anche disponibili su alcune calcolatrici).

Calcolo delle aree sottese alla curva

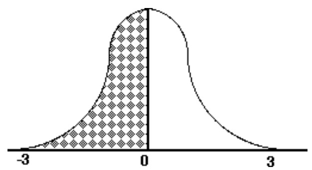
Esempio_1: calcolare l'area compresa fra -∞ e 1.83, in simboli: $\phi(1.83)$



$\phi(1.83) = 0.9664$

$P(Z \le 1.83)$

Esempio_2: calcolare l'area compresa fra -∞ e 0, in simboli: $\phi(0)$

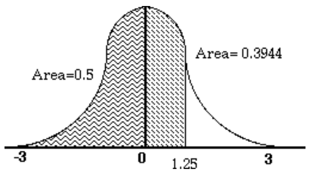


$2\phi(0)=1$ e quindi $\phi(0)=\frac{1}{2}$

$P(Z \le 0)$

Altri esempi

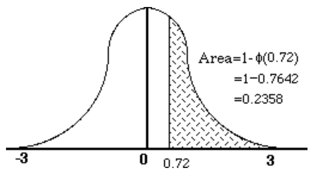
Esempio_3: calcolo dell'area tra -∞ e 1.25 $P(Z \le 1.25)$



Alcune tavole danno invece l'area compresa tra il valore dato e lo zero. Per ottenere l'intera area occorre sommare 0.5 che è l'area a sinistra dello zero.

$\phi(1.25)$ =0.5 + 0.3944 = 0.8944

Esempio_4: calcolo dell'area a destra di 0.72 $P(Z > 0.72)$



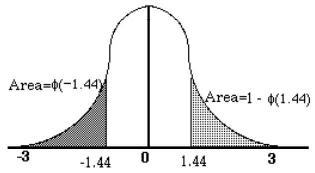
Poiché l'area totale è uguale ad uno si ha l'identità:

Area a sinistra=1- Area a destra

$$1 - P(Z \le 0.72)$$

Ancora esempi

Esempio_5: calcolo dell'area per valori negativi: ϕ(-1.44) $P(Z \le -1.44)$



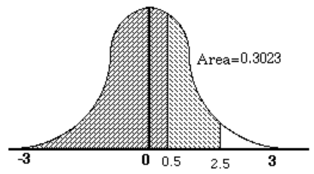
La curva normale è simmetrica per cui

$$\phi(-X) = 1 - \phi(X)$$

Per calcolare l'area a sinistra di un valore negativo si usa l'area a destra del valore positivo corrispondente:

ϕ(-1.44) =1-ϕ(1.44) = 1- 0.9251 = 0.0749

Esempio_6: calcolo dell'area compresa tra due valori positivi. $P(0.5 \le Z \le 2.5)$



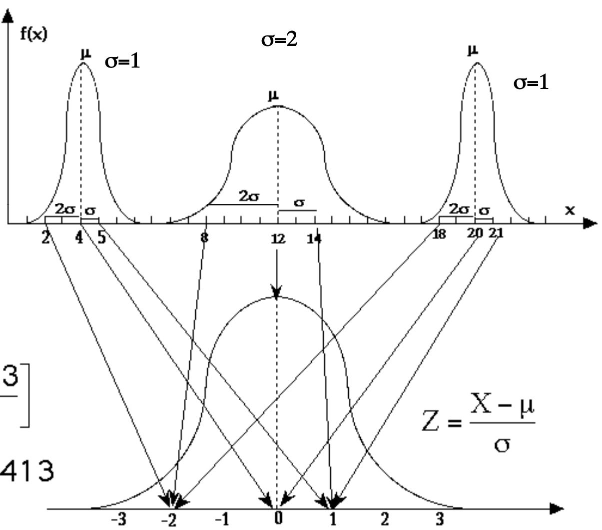
L'area compresa tra due valori qualsiasi si ottiene per differenza tra le aree a sinistra degli estremi dell'intervallo

Area=ϕ(2.5) - ϕ(0.5) =0.9938 - 0.6915 =0.3023

Generalizzazione

Le tre diverse v.c. normali possono essere ricondotte ad una sola distribuzione con la trasformazione in unità standard.

Le aree sottese a X~(μ,σ) sono identiche a quelle dellaZ~N(0,1) dopo aver trasformato la X in unità standard



$$X \sim N(3,4)$$

$$P(X < 5) = P\left[\frac{X-3}{2} \le \frac{5-3}{2}\right]$$

$$= P(Z \le 1) = 0.8413$$

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

Esempio

Sia X una v.c. Normale con media μ=16 e deviazione standard σ=5.

1) Calcolare: $P(11 < X < 21)$

$$P(11 < X < 21) = P\left(\frac{11-16}{5} < \frac{X-16}{5} < \frac{21-16}{5}\right) = P(-1 < Z < 1)$$

$$= \phi(1) - \phi(-1) = \phi(1) - [1 - \phi(1)] = 2\phi(1) - 1$$

$$= 2 * 0.8413 - 1 = 0.6826$$

2) Calcolare: $P(X > 26)$

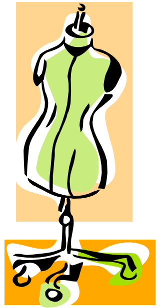
$$P(X > 26) = 1 - P(X \le 26) = 1 - P\left(\frac{X-16}{5} \le \frac{26-16}{5}\right) = 1 - P(Z \le 2)$$

$$= 1 - 0.9772 = 0.0228$$

Esempio

Supponiamo che le vendite di vestiti di un negozio specializzato seguano la distribuzione Normale con una media di 36 vendite al giorno ed una deviazione standard di 9 vendite al giorno.

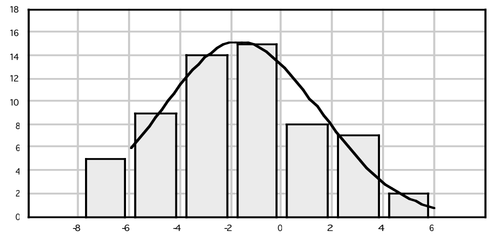
Calcolare la probabilità che in un dato giorno si vendano più di 12 vestiti



$$P(X > 12) = 1 - P(X \le 12) = 1 - P\left(\frac{X-36}{9} \le \frac{12-36}{9}\right) = 1 - P(Z \le -2.67)$$
$$= P(Z \le 2.67) = \phi(2.67) = 0.9962$$

Uso della normale/1

Si è rilevato l'errore (rispetto ad un valore prefissato) nel peso di un campione di 60 lotti.



-7.2	-4.8	-1.7	-3.1	-0.7	1.6
-6.7	-4.8	-1.6	-3.1	-0.5	2.2
-6.7	-4.8	-1.5	-2.9	-0.1	2.3
-6.6	-4.6	-1.4	-2.8	0.3	3.4
-6.5	-3.8	-1.4	-2.7	0.3	3.4
-5.6	-3.7	-1.3	-2.7	0.6	3.5
-5.5	-3.6	-1.3	-2.2	0.8	3.5
-5.3	-3.5	-1.2	-2.2	1.3	3.6
-5.1	-3.5	-1.0	-1.9	1.5	4.2
-5.0	-3.3	-0.8	-1.8	1.5	5.5

$$\hat{\mu} = -1.683, \quad s = 3.106$$

Ipotizziamo che tali valori siano validi per tutti i possibili lotti.

Se si sceglie a caso un lotto, qual è la probabilità che l'errore nel peso sia compreso tra -0.5 e +0.5

$$P(-0.5 \le x \le 0.5) = P\left(\frac{-0.5+1.683}{3.106} \le z \le \frac{0.5+1.683}{3.106}\right) = P(0.38 \le z \le 0.70)$$
$$= \Phi(0.7) - \Phi(0.38) = 0.758 - 0.648 = 0.11 \Rightarrow 11\%$$

Approssimazione di v.c. Discrete

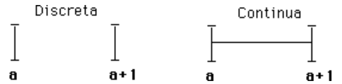
In alcuni casi e nonostante la contraddizione, le distribuzioni delle v.c. discrete possono essere approssimate dalla v.c. normale o gaussiana

Occorre però tenere conto che i valori possibili delle discrete sono assimilabili ad un insieme di numeri interi naturali: 0, 1, 2, ..., n,... laddove la Gaussiana può assumere ogni valore reale

Nel discreto si ha $P(X \le a) + P(X \ge a+1) = 1$

Nel continuo $P(X \le a) + P(X \ge a+1) \le 1$

in quanto non sono inclusi i valori dell'intervallo $a < X < a+1$



Approssimazione di v.c. discrete/2

Per superare tali problemi si usa un fattore di correzione della continuità pari a 0.5.

Se "X" è discreta e deve essere approssimata con una continua "Y" si ha:

$$P(X = a) \cong P(a - 0.5 \le Y \le a + 0.5)$$

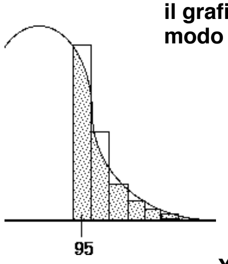
Inoltre, per recuperare la differenza tra includere o no l'estremo nelle discrete, si pone

Sono convenzioni

$$P(X < a) = P(Y \le a - 0.5); \qquad P(X \le a) = P(Y \le a + 0.5)$$
$$P(X > a) = P(Y > a + 0.5); \qquad P(X \ge a) = P(Y \ge a - 0.5)$$

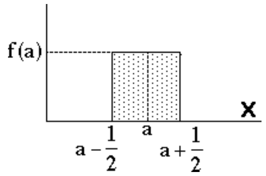
Esempio

Le risposte esatte "X" in un test attitudinale hanno μ=80 e σ=10. Calcolare, con la gaussiana "Y", la frazione di candidati con punteggio inferiore a 95



il grafico della f(x) assume che gli istogrammi abbiano base "1" in modo che l'area sia uguale alla probabilità.

$$P(X = a) = f(a)*1$$



La normale si usa per approssimare le aree degli istogrammi

X ≥ 95 significa in realtà ≥ 94.5
X ≤ 95 significa in realtà ≤ 95.5

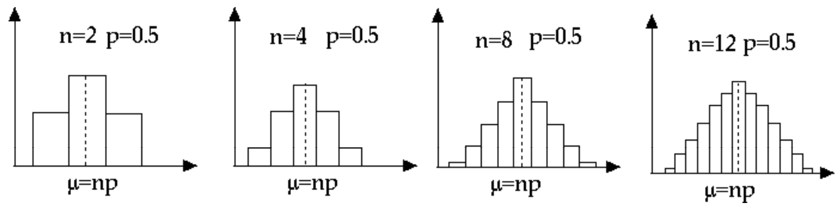
Calcolo relativo all'esempio:

$$P(X < 95) \cong P(Y \le 94.5) = P\left(\frac{Y-80}{10} \le \frac{94.5-80}{10}\right) = \phi(1.45) = 0.9265$$

Approssimazione della Binomiale

La Normale è una buona approssimazione della Binomiale se

- **"n", il numero delle prove, è molto grande**
- **"p", la probabilità di successo, è molto vicino a 0.5**



Come si vede, all'aumentare di "n" la distribuzione assume una forma sempre più campanulare giustificando l'uso della normale.

Regola pratica

la procedura per l'approssimazione della binomiale con la normale ha 3 passi:

1. **Si calcolano** $\mu = n*p; \quad \sigma = \sqrt{n*p*(1-p)}$

2. **Si applica il fattore di correzione della continuità per convertire la v.c. discreta in continua**

$$Z = \frac{(X-0.5)-\mu}{\sigma}; \quad \text{oppure} \quad Z = \frac{(X+0.5)-\mu}{\sigma}$$

3. **Si usano le tavole della normale standardizzata per determinare le probabilità della binomiale**

$$P(X = a) \cong \phi\left(\frac{a+0.5-\mu}{\sigma}\right) - \phi\left(\frac{a-0.5-\mu}{\sigma}\right)$$

$$P(X < b) \cong \phi\left(\frac{b-0.5-\mu}{\sigma}\right)$$

$$P(X \le b) \cong \phi\left(\frac{b+0.5-\mu}{\sigma}\right)$$

Esempio

Le nuove imprese hanno il 50% di probabilità di fallire al primo anno. Usando la formula della binomiale, la probabilità che 40 (su n=100) imprese siano ancora in vita dopo il primo esercizio è

$$P(X = 40) = \frac{100!}{40!60!}(0.5)^{40}(0.5)^{60} = 0.0108$$

Usiamo l'approssimazione normale

1. $\mu = n*p = 100*0.5 = 50; \quad \sigma = \sqrt{50*0.5} = 5$

2. $P(X = 40) \cong P(39.5 \le X \le 40.5) = P\left(\frac{39.5-50}{5} \le Z \le \frac{40.5-50}{5}\right)$
$= P(-2.1 < Z < -1.9)$

3. $\phi(-1.9) - \phi(-2.1) = 1 - \phi(1.9) - [1 - \phi(2.1)] = 0.9821 - 0.9713 = 0.0108$

Esempio

il 20% della "Gente" è convinta di sopravvivere ad un conflitto nucleare. Si indichi con "Y" il numero di tali persone in un campione di ampiezza n=25.

a) Si adoperi l'approssimazione Normale della Binomiale per determinare: $P(6 \le Y \le 9)$
b) La si confronti con il valore ottenuto usando direttamente la Binomiale.

a) $\hat{\mu} = 25 * 0.20 = 5; \quad \hat{\sigma} = \sqrt{\hat{\mu} * 0.8} = 2$

$$P(6 \le Y \le 9) = P(5.5 \le X \le 9.5) = P\left(\frac{5.5-5}{2} \le \frac{X-5}{2} \le \frac{9.5-5}{2}\right)$$
$$= P(0.25 \le Z \le 2.25) = \phi(2.25) - \phi(0.25) = 0.9878 - 0.5987$$
$$= 0.3891$$

b) $P(6 \le Y \le 9) = \sum_{i=0}^{9}\binom{25}{i}0.2^i 0.8^{25-i} - \sum_{i=0}^{6}\binom{25}{i}0.2^i 0.8^{25-i} = 0.9827 - 0.7800 = 0.2027$

L'asimmetria della binomiale (dovuta al parametro "p" lontano da 0.5) ha comportato una notevole differenza nell'approssimazione della probabilità

Approssimazione normale della Poisson

Anche la Poisson è approssimabile dalla normale ponendo $\mu = \lambda; \quad \sigma = \sqrt{\lambda}$

però si deve avere una media elevata, diciamo, $\lambda \ge 10$

ESEMPIO

il numero delle vendite settimanali di un prodotto è ben approssimato dalla Poisson con una media di 2 vendite al giorno. Calcolare la probabilità che in una settimana si vendano meno di 23 prodotti

Dato che in media si vendono 2 unità al giorno, in media, si venderanno 14 prodotti in una settimana, per cui

$$P(0 \le X \le 23) = \sum_{i=0}^{23}\frac{(14)^i e^{-14}}{i!} = 0.99067$$

Approssimazione Normale con $\mu = \lambda = 14; \quad \sigma = \sqrt{14} = 3.7417$

$$P(0 \le X \le 23) \cong \phi\left(\frac{23+0.5-14}{3.7417}\right) - \phi\left(\frac{0-0.5-14}{3.7417}\right) = \phi(2.54) - \phi(-3.88) =$$
$$= 0.9944$$

La curtosi

La discussione è limitata alle curve unimodali e simmetriche di variabili continue

Curtosi è la deformazione cui deve essere sottoposta una curva normale, di pari media e pari deviazione standard, per sovrapporsi alla curva data

- **Curva Leptocurtica** — **Curva che presenta, rispetto alla normale, ordinate più alte al centro (la sommità è più acuminata) e code spesse e allungate**
- **Curva Platicurtica** — **Curva che ha ordinate più alte nei valori intermedi (la sommità è più appiattita) e code sottili e brevi.**
- **Curva Mesocurtica** — **Curva con sommità e code simili a quelle della normale**

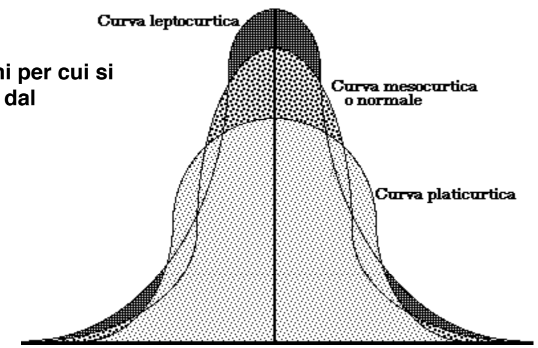
Considerazioni sulla curtosi

La curtosi è considerata una disnormalità.

Come nell'asimmetria è importante accertare il grado di allontanamento dal modello normale.

E' anche importante capire le ragioni per cui si verifica l'eventuale scostamento dal modello di riferimento

La curtosi è un concetto composito: le code e la sommità andrebbero studiate separatamente.



Misura della curtosi

La curtosi è misurata con il momento quarto della standardizzata

$$m(4,0)=\sum_{i=1}^{k}\left[\frac{C_i - \mu}{\sigma}\right]^4 f_i; \quad = \sum_{i=1}^{k} z_i^4 f_i$$

Poiché per la normale standardizzata si ha $\mu(4,0)=3$, si adopera, come indice di curtosi, la misura

$$\gamma 2 = \sum_{i=1}^{k} z_i^4 f_i - 3$$

- Per distribuzioni platicurtiche l'indice $\gamma_2<0$
- Per distribuzioni leptocurtiche $\gamma_2>0$
- Per distribuzioni mesocurtiche $\gamma_2=0$

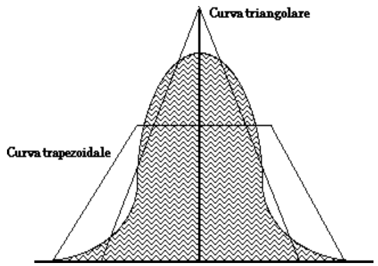
L'indice è legato a quello di asimmetria dalla relazione: $\gamma_2 - (\gamma_1)^2 \geq 1$

che può essere utile nel controllo dei calcoli

Difficoltà nel concetto di curtosi

Il problema con la curtosi è che la disnormalità è dovuta a due fattori distinti:

GRADO DI APPUNTIMENTO AL CENTRO SPESSORE DELLE CODE



Per raggiungere lo stesso grado di appuntimento della curva triangolare, la normale dovrà essere tirata verso l'alto e assottigliata nei fianchi senza modificare le code.

Per portare le code della normale allo spessore della curva trapezoidale, si possono allargare i fianchi senza intervenire sulla sommità.

L'effetto congiunto di questi due fattori porta a valori di γ_2 incoerenti con l'aspetto della distribuzione di frequenza

L'indice di Groeneveld e Meeden

Per ovviare alla mancanza di univocità tra i valori di $\gamma 2$ e la **morfologia** delle curve ci sono diversi indici

In particolare c'è quello suggerito dai suddetti autori

$$GM=\frac{(X_{0.85} - Q_{0.75}) - (Q_{0.75} - X_{0.65})}{(X_{0.85} - Q_{0.75}) + (Q_{0.75} - X_{0.65})} = \frac{X_{0.85} + X_{0.65} - 2Q_{0.75}}{X_{0.85} - X_{0.65}}$$

L' indice varia nell'intervallo [-1,1] ed è invariante rispetto a trasformazioni lineari.

Valori grandi e positivi di GM saranno associati a curve leptocurtiche laddove curve platicurtiche daranno luogo a valori positivi ma piccoli (per la normale standardizzata si ha GM=0.073).

Il valore nullo di GM si ha solo per distribuzioni uniformi ed un GM negativo si avrà per curve a forma di U

Trasformazioni per la normalità

Le indagini statistiche coinvolgono solitamente moltissime variabili e non tutte possono essere indagate per gli aspetti finora considerati: centralità, variabilità e forma.

In genere ci si limita ai primi due aspetti e per il terzo si assume che la forma della poligonale delle frequenze sia simile alla curva normale. In moltissimi casi tale assunzione regge. Altre volte è meno valida: ad esempio nelle distribuzioni ad "U", a "J"

Le trasformazioni per la normalità sono delle modifiche applicate alle modalità originarie tali che la loro poligonale è più simile alla normale.

La procedura di scelta della trasformazione può essere automatizzata

La trasformazione di Box-Cox

Fra le molte trasformazioni possibili esamineremo la più famosa

Per $X_i > \mathbf{0}$

$$Y_i = f(X_i;\lambda) = \begin{cases} \dfrac{X_i^{\lambda} - 1}{\lambda} & \text{per } \lambda \neq 0 \\ Ln(X_i) & \text{per } \lambda = 0 \end{cases}$$

Esempi.

Per $\lambda=1$ si ha la trasformazione lineare: $Y_i=X_i-1$

Per $\lambda=\frac{1}{2}$ si ha la trasformazione radicale: $Y_i= 2(\sqrt{X_i}-1)$

Per $\lambda=-1$ si ha la trasformazione reciproca: $Y_i= 1 - \frac{1}{X_i}$

- **Di solito il parametro λ è incognito e per stimarlo occorre fare ricorso ad una routine di calcolo.**
- **Un modo meno rigoroso, ma abbastanza efficace è quello di calcolare i due coefficienti $\gamma 1$ e $\gamma 2$ (od anche gli analoghi indici basati sui percentili) per alcuni valori predefiniti di λ**

{-2.5, -2.0, - 1.5, -1, -0.5, 0, 0.5, 1.0, 1.5, 2.0 2.5}

Modelli di densità

L' interesse per i modelli è duplice:

Per il loro carattere descrittivo che raggiunge l'obiettivo con la comoda scorciatoia dell'interpolazione del poligono delle frequenze.

I calcoli sono più rapidi, meglio approssimati e più resistenti agli errori accidentali.“...

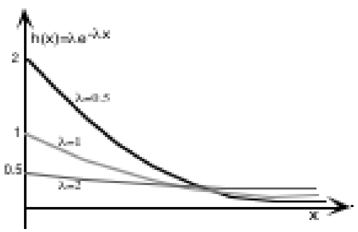
Talvolta nello studio dei dati reali è diffficile ottenere distribuzioni i cui grafici presentino andamenti così regolari da ricordare quelli di funzioni matematiche ben note.

In genere bisogna determinare un modello esplicativo che si adatti bene alla distribuzione osservata scegliend fra piuù modelli alternativi

Esempio

Se una sequenza di eventi Y si manifesta nel tempo secondo i modello di Poisson, quanto bisogna aspettare perché si verifichi il primo?

Sia λ il numero medio di “eventi” e X = “Tempo antecedente la prima manifestazione”.
L'evento X>x si realizza se nel lasso temporale [0,x] la Poisson non ha prodotto successi, cioè Y=0. Quindi:



$$P(X > x) = P(Y = 0) = \frac{(\lambda x)^0 e^{-\lambda x}}{0!} = e^{-\lambda x} \Rightarrow P(X \le x) = F(x) = 1 - e^{-\lambda x}$$

Questo modello ha numerose applicazioni nei tempi di attesa in coda.

La funzione di densità si ricava con il calcolo differenziale:

$$F'(x) = h(x) = \begin{cases} \lambda e^{-\lambda x} & per \quad x > 0 \\ 0 & altrove \end{cases}; \quad \lambda > 0$$

Funzione di rischio

L'inaffidabilità ha come indicatore la frequenza con la quale si succedono nel tempo le disfunzioni.

I modelli di variabili casuali descrivono la tenuta di un sistema

La scelta del modello è facilitata dalla funzione di rischio

$$r(x) = -\frac{d}{dx} Log[1 - F(x)] = \frac{h(x)}{1 - F(x)} = \lim_{\Delta t \to 0} \frac{P(t \le x \le t + \Delta t)}{\Delta t}$$

r(x) esprime la probabilità di errore a x+Dx dopo che l'elemento è rimasto operativo fino a “x”.

La sorte provoca rotture rapide in sistemi con tassi di rischio basso e durate inusitate per apparati di dubbia affidabilità

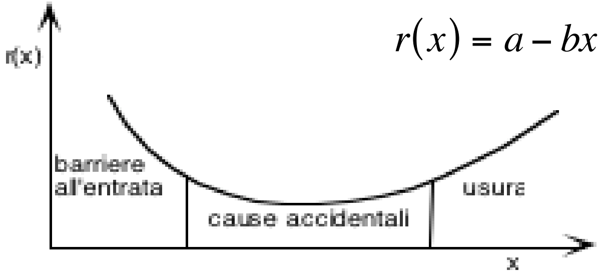
Esempio: curva ad “U”

Nel primo tratto il funzionamento corre pericoli soprattutto a causa delle barriere all’entrata: cattive condizioni iniziali, difetti all’origine, eventi eccezionali;

Nella fase centrale dominano i fattori accidentali

Nella terza fase le cause più verosimili sono l’usura e il deterioramento delle condizioni di operatività

$$r(x) = a - bx + cx^2; \; a,b,c > 0$$



Dal rischio alla densità

Se è nota la funzione di rischio è possibile ricavare la funzione di densità e di ripartizione.

$$\int_0^x r(t)dt = -Ln[1 - F(x)] \Rightarrow F(x) = 1 - e^{-\int_0^x r(t)dt}, \quad h(x) = r(x)e^{-\int_0^x r(t)dt}$$

Quadratica: $$h(x) = \left[a - bx + cx^2\right] e^{-\left[ax - \frac{b}{2}x^2 + \frac{c}{3}x^3\right]}$$

Costante: $$r(x) = \lambda \Rightarrow h(x) = \lambda e^{-\int_0^x \lambda dt} = \lambda e^{-\lambda t|_0^x} = \lambda e^{-\lambda x}$$

Var. cas. Assolutamente continue

Una funzione reale F(.) definita in R è assolutamente continua se

$$x_1 < x_2 \Rightarrow F(x_2) - F(x_1) = \int_{x1}^{x2} h(t)dt$$

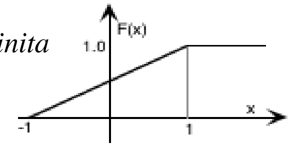
cioè se F(.) è l’integrale definito della sua funzione di densità:

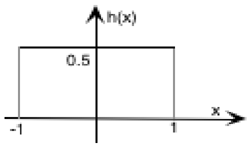
$$F(x) = \int_{-\infty}^{x} h(t)dt$$

Controesempio

$$F(x) = \begin{cases} 0 & se \quad x \le -1 \\ \frac{x+1}{2} & se \quad -1 < x \le 1 \\ 1 & se \quad x > 1 \end{cases} \Rightarrow h(x) = F'(x) = \begin{cases} 0 & se \quad x < -1, \; x > 1 \\ \frac{1}{2} & se \quad -1 < x < 1 \end{cases}$$

La derivata non è definita Negli estremi

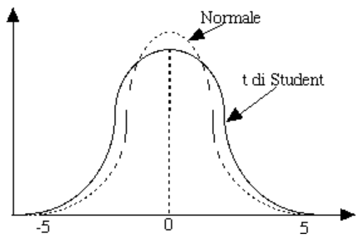




La t di Student

Questo modello è molto simile a quello gaussiano, ma ha code più ”pesanti” (ordinate estreme più alte)

GUINNESS



$$-\infty < t_n < \infty, \quad n > 2$$

$$E(t_n) = 0, \quad Var(t_n) = \frac{n}{n-2}$$

La varianza è superiore all'unità, ma si avvicina ad uno all'aumentare di "n"

L'elemento caratterizzante della t di Student sono "i gradi di libertà" n che è il parametro della t di Student.

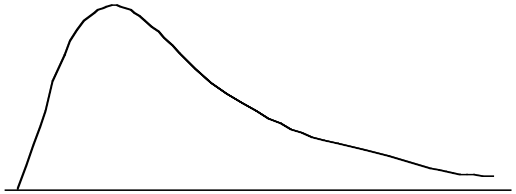
Per ogni grado di libertà esiste una t di Student, sebbene queste diventino poco distinguibili per n≥60.

Questa v.c. è stata analizzata da W.S. Gosset, nel 1906, che firmò l'articolo con lo pseudonimo "STUDENT" ed è da allora nota come "La t di Student"

χ^2 (chi-quadrato)

Questo modello è definito per i soli non negativi e presenta una marcata asimmetria positiva

Anche in questo caso l'elemento caratterizzante sono "i gradi di libertà" cioè "g"

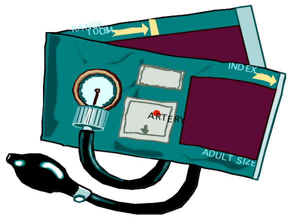


$$E(\chi^2) = g, \qquad Var(\chi^2) = 2g$$

Per "g" superiore a 30 la distribuzione del χ^2 si avvicina a quella normale

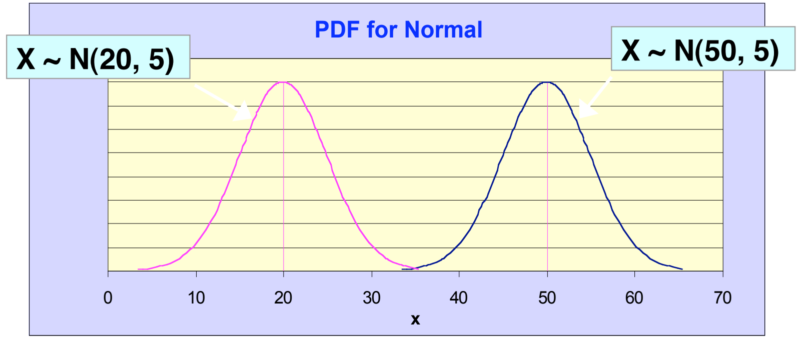
La distribuzione del χ^2 si incontra nello studio dell'adattamento e della variabilità

Questa è importante nelle analisi cliniche e nel controllo della qualità



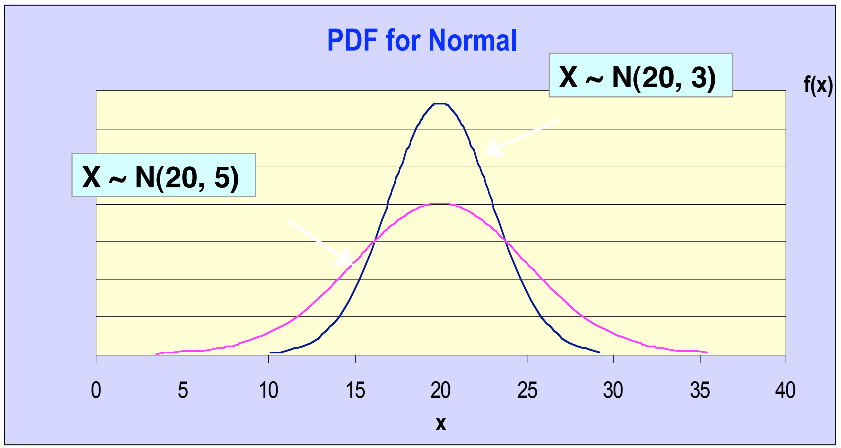
Parametri delle distribuzioni

1. **Posizione: specificano la tendenza centrale dei valori nel modello od anche il punto di massimo addensamento**



Parametri delle distribuzioni/2

2. **Scala: determinano l'unità di misura della variabile (espansione o contrazione della scala delle ascisse)**



Parametri delle distribuzioni/3

3. Forma: Controllo la forma che assume il modello di probabilità. In particolare, l'asimmetria, l'appuntimento al centro, la pesantezza delle code