

VARIABILI ALEATORIE CONTINUE

Esistono parecchi fenomeni reali per la cui descrizione le variabili aleatorie discrete non sono adatte. Per esempio è necessaria una variabile aleatoria *continua* (ovvero una variabile che può assumere tutti i valori reali in un dato intervallo), per descrivere il tempo di vita di un apparecchio soggetto a guasti casuali, o per descrivere il gioco del lancio delle freccette. Inoltre facendo ad esempio riferimento al lancio delle freccette, è inutile chiedersi la probabilità che la freccetta colpisca esattamente un punto sul disco, ma ha più senso cercare di capire la probabilità che la freccetta colpisca una zona ben definita del disco. Si è dunque interessati alla probabilità non di eventi rappresentati da singoli punti ma di eventi rappresentati da intervalli. Proprio per questo una variabile aleatoria X è definita assegnando una funzione che permette di calcolare la probabilità che X assumi valori in un qualsivoglia intervallo.

Definizione Una variabile aleatoria continua X , viene definita assegnando una funzione f , detta *densità (di probabilità)* di X , che soddisfi le seguenti proprietà:

- $f(x) \geq 0$ per ogni $x \in \mathbb{R}$,
- $\int_{-\infty}^{+\infty} f(x)dx = 1$

La probabilità che X assuma valori in un determinato intervallo I è data dall'integrale della sua densità sull'intervallo I .

Osservazioni:

- a) Se l'intervallo $I = [a, b]$ si riduce ad un punto, cioè se $a = b$, risulta che:

$$p(X \in I) = \int_a^b f(x)dx \quad e \quad p(X = a) = \int_a^a f(x)dx = 0$$

Perciò se X è una variabile aleatoria continua, la probabilità che essa assuma un qualsivoglia valore reale prefissato è sempre nulla.

- b) Data la densità di probabilità f di una variabile aleatoria continua X , il valore $f(a)$ da essa assunto quando $x = a$, NON ha (come accade nel caso discreto) il significato di probabilità dell'evento $X = a$, infatti questa probabilità è sempre uguale a zero, mentre il valore assunto da f in a , in generale è un numero positivo.

Diamo ora alcune definizioni che riprendiamo dal caso discreto:

Definizioni: Data una variabile aleatoria continua X , di densità f ,

- si dice *media* o *valore atteso* di X e si indica con il simbolo $E(X)$ (o con la lettera μ), il numero se esiste, così definito

$$\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

- si dice *varianza* di X e si indica con il simbolo $V(X)$, o con σ^2 , il numero, se esiste, così definito

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

- si definisce *deviazione standard* di X (e si indica con il simbolo σ) la radice quadrata della varianza

$$\sigma = \sqrt{V(x)}$$

Definizione Sia una variabile aleatoria continua X , avente come densità la funzione f , si chiama *funzione di ripartizione* di X la funzione che, per ogni $x \in \mathbb{R}$, è così definita:

$$F(x) = p(X \leq x) = \int_{-\infty}^x f(t)dt$$

Esistono diverse funzioni notevoli di densità di probabilità :

- uniforme,
- binomiale
- Poisson
- Gauss,

Concentreremo la nostra attenzione in modo particolare sull'ultima.

8.3 Distribuzione normale (o di Gauss)

La funzione di densità di questa particolare distribuzione è data da:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \forall x \in \mathbb{R}$$

Con μ e σ rispettivamente valor medio e deviazione standard di una variabile aleatoria continua (v.a.c.) X . Cerchiamo di capire l'andamento di $f(x)$ e andiamo a studiarla.

- Dominio $\mathcal{D} = \mathbb{R}$
- Segno $f(x) > 0 \quad \forall x \in \mathbb{R}$, è simmetrica rispetto alla retta $x = \mu$.
- Limiti agli estremi del dominio:

$$\lim_{x \rightarrow \pm\infty} f(x) = \lim_{t \rightarrow \pm\infty} e^{-t^2} = 0$$

La retta $y = 0$ è asintoto orizzontale.

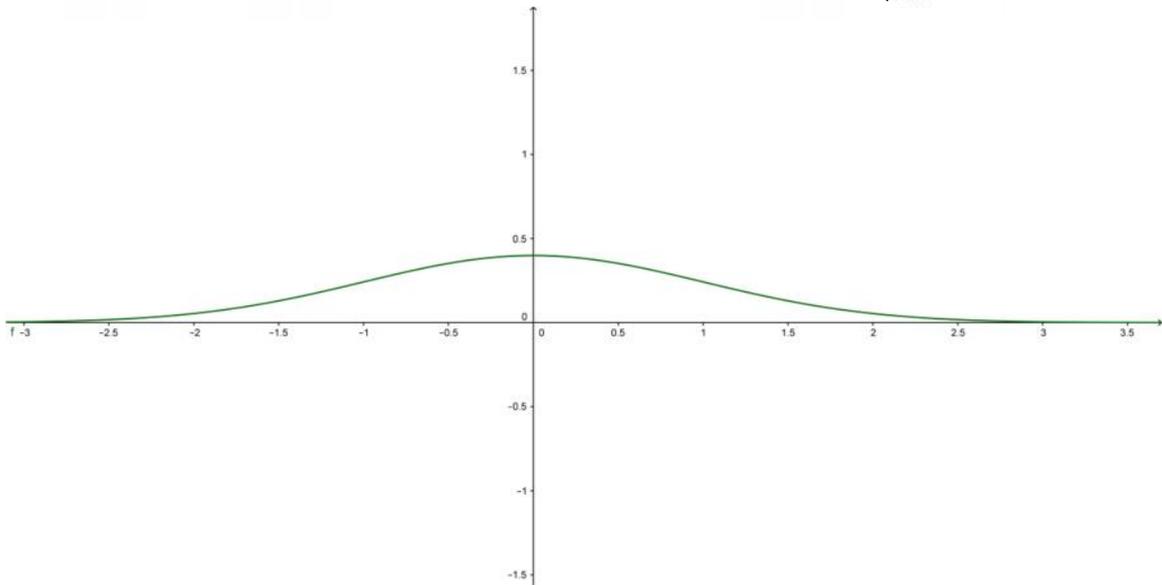
- Crescenza e decrescenza: $f'(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot \left(-\frac{2}{2\sigma^2}(x-\mu)\right) \geq 0 \Leftrightarrow x \leq \mu$

Si ha che $x = \mu$ è punto di massimo, e il massimo vale $f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$

- Concavità e convessità: $f''(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot \left(-\frac{1}{\sigma^2} + \frac{(x-\mu)^2}{\sigma^4}\right) \geq 0$

Se e solo se $\frac{(x-\mu)^2}{\sigma^2} - 1 \geq 0 \Leftrightarrow (x-\mu)^2 \geq \sigma^2 \Leftrightarrow x \leq -\sigma + \mu \wedge x \geq \sigma + \mu$

$x = \mu - \sigma$ e $x = \mu + \sigma$ sono punti di flesso. Il flesso vale: $f(\mu + \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}}$.



Esempi di caratteristiche che si distribuiscono normalmente.

- i. Errori di misurazione di una grandezza fisica
- ii. Peso e altezza di una popolazione omogenea.
- iii. Dimensioni di oggetti prodotti in serie.

La probabilità $p(X \leq x) = F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ si calcola utilizzando varie tabelle dove è riportato il valore della DISTRIBUZIONI NORMALE STANDARD.

Sia X una v.a.c. con media μ e scarto quadratico σ , allora diremo che $X \in \mathcal{N}(\mu, \sigma)$.

Definizione Sia X una v.a.c. con $X \in \mathcal{N}(0,1)$, funzione di densità $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, la funzione di partizione $F(x) = p(X \leq x) = \int_{-\infty}^x f(t)dt$ associata ad X è detta *distribuzione normale standard*.

$f(x)$ è una funzione pari e solitamente si indica con $\varphi(x) = p(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$.

Tabella 1. Funzione di ripartizione della distribuzione normale standard

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Esempi:

$$\Phi(0.05) \simeq 0.5199, \Phi(1.28) \simeq 0.8997, \Phi(3.51) \simeq 1.0000$$

$$\Phi(-2.41) = 1 - \Phi(2.41) \simeq 1 - 0.9920 = 0.0080, \Phi(-4.03) = 1 - \Phi(4.03) \simeq 1 - 1 = 0.0000$$

$$\Phi(x) = 0.651 \implies x \simeq 0.39, \Phi(x) = 0.9921 \implies x \simeq 2.415,$$

$$\Phi(x) = 0.374 \implies 1 - \Phi(-x) = 0.374 \implies \Phi(-x) = 1 - 0.374 = 0.626 \implies -x \simeq 0.32 \implies x \simeq -0.32$$

REGOLE GENERALE:

- i. $p(X \leq x) = \varphi(x)$
- ii. $p(X > x) = 1 - \varphi(x)$
- iii. $p(X \leq -x) = \varphi(-x) = 1 - \varphi(x)$
- iv. $p(-x \leq X \leq x) = \varphi(x) - \varphi(-x) = \varphi(x) - (1 - \varphi(x)) = 2\varphi(x) - 1$
 $= 1 - 2\varphi(-x) = 1 - 2(1 - \varphi(x)) = 1 - 2 + 2\varphi(x) = 2\varphi(x) - 1$

Dalle tavole si ottiene che :

- $p(X \leq 0) = \frac{1}{2}$
- $p(-1 \leq X \leq 1) = 2\varphi(1) - 1 \cong 0.6827$
- $p(-2 \leq X \leq 2) \cong 0.9545$
- $p(-3 \leq X \leq 3) \cong 0.9973$

Come facciamo però se X non è standard? Dobbiamo trovare un modo per renderla standard!!!

STANDARIZZAZIONE DI UNA VARIABILE ALEATORIA CONTINUA

Se X è una variabile aleatoria continua con $X \in \mathcal{N}(\mu, \sigma)$, allora $Z = \frac{X-\mu}{\sigma}$ è una v.a.c. con $Z \in \mathcal{N}(0,1)$ che si chiama *standarizzazione di X*.

Dunque:

$$p(X \leq x) = p\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = p\left(Z \leq \frac{x-\mu}{\sigma}\right) = \varphi\left(\frac{x-\mu}{\sigma}\right)$$

Quindi qualsiasi sia la v.a.c. che considero, grazie alla standarizzazione riesco a leggere la probabilità che mi interessa.

Da quanto detto precedentemente

- $p(\mu - \sigma \leq X \leq \mu + \sigma) = p(-1 \leq Z \leq 1) \cong 0.6827$
- $p(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = p(-2 \leq Z \leq 2) \cong 0.9545$
- $p(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = p(-3 \leq Z \leq 3) \cong 0.9973$

Esercizio (uso della tabella)

Il livello di glucosio a digiuno in un gruppo di pazienti non diabetici si distribuisce normalmente con $\mu = 98 \text{ mg}/100 \text{ ml}$ e $\sigma = 8 \text{ mg}/100 \text{ ml}$.

Determinare:

- a) $p(X < 82 \text{ mg}/100 \text{ ml})$
- b) $p(90 \leq X \leq 106 \text{ mg}/100 \text{ ml})$
- c) $p(X > 116 \text{ mg}/100 \text{ ml})$

SVOLGIMENTO

$X \in \mathcal{N}(98,8)$ utilizzando la sua standardizzazione Z otteniamo:

- a) $p(X < 82 \text{ mg}/100 \text{ ml}) = p\left(Z < \frac{82-98}{8}\right) = p(Z < -2) = \varphi(-2) = 1 - \varphi(2) = 1 - 0.9772 = 0.0228 = 2,28 \%$
- b) $p(90 \leq X \leq 106) = p\left(\frac{90-98}{8} \leq Z \leq \frac{106-98}{8}\right) = p(-1 \leq Z \leq 1) \cong 68,26\%$
- c) $p(X > 116) = p\left(Z > \frac{116-98}{8}\right) = p(Z > 2,25) = 1 - p(Z < 2,25) = 1 - \varphi(2,25) = 1 - 0,9878 \cong 1,22\%$.

Teorema (del limite centrale)

Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti, aventi tutte la stessa distribuzione di probabilità, (quindi stessa media μ e stessa varianza σ^2). Allora

$$\lim_{n \rightarrow +\infty} p\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x\right) = p(\mathcal{N}(0,1) \leq x)$$

Qual è il suo significato? In sostanza possiamo approssimare la media di X_1, X_2, \dots, X_n con una variabile normale avente la media comune μ e la varianza $\frac{\sigma^2}{n}$.

Questo teorema ci porta a formulare delle stime attraverso gli intervalli di confidenza.

INTERVALLI DI CONFIDENZA.

Sia X una caratteristica di una popolazione P che vogliamo studiare. Supponiamo che X sia normalmente distribuita con media μ e varianza σ^2 , di cui non siamo sempre a conoscenza. Estraiamo un campione casuale di n elementi di P , siano x_1, x_2, \dots, x_n i valori di X per gli elementi del campione.

Possiamo pensare di avere X_1, X_2, \dots, X_n v.a.c. del tipo $\mathcal{N}(\mu, \sigma^2)$ e che x_1, x_2, \dots, x_n siano i valori assunti da X_1, X_2, \dots, X_n dopo aver svolto l'esperimento.

Sia $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ la media campionaria.

Per il teorema del limite centrale $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0,1)$,

ossia per $n \gg 0$ si ha che $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

Possiamo usare queste informazioni per ottenere stime di tipo probabilistico per la media μ :

Scegliamo una probabilità $P \in (0,1)$, possiamo costruire un intervallo (μ_1, μ_2) , detto intervallo di confidenza tale che $p(\mu \in (\mu_1, \mu_2)) = P$.

Esempio Sia X una v.a.c con $\mathcal{N}(\mu, \sigma^2)$ con $\sigma^2 = 9$. Da un campione casuale di 5 elementi della popolazione P si ottiene $\bar{x} = 61$.

Determinare intervallo (μ_1, μ_2) tale che la $p(\mu \in (\mu_1, \mu_2)) = 95\%$

Lo scopo è quindi quello di determinare un intervallo di confidenza al 95% per μ .

Sto quindi cercando Z una v.a.c. del tipo $\mathcal{N}(0,1)$ tale che $p(-z \leq Z \leq z) = 0,95$. Determinato Z ,

finalmente posso trovare la media μ sfruttando la relazione $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$.

Ora:

$p(-z \leq Z \leq z) = 0,95$ ma $p(-z \leq Z \leq z) = 2p(Z < z) - 1$ quindi

$$2p(Z < z) - 1 = 0,95 \Rightarrow p(Z < z) = \frac{0,95 + 1}{2} = 0,9750$$

Cerco 0,9750 nella tabella e vedo che corrisponde a $z = 1,96$.

Otteniamo dunque $-1,96 < z < 1,96$ che per il teorema del limite centrale

$$-1,96 < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1,96$$

Sviluppando i conti:

$$\begin{aligned}
-1,96 \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1,96 \cdot \frac{\sigma}{\sqrt{n}} \\
-1,96 \cdot \frac{\sigma}{\sqrt{n}} - \bar{X} < -\mu < 1,96 \cdot \frac{\sigma}{\sqrt{n}} - \bar{X} \\
\bar{X} - 1,96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1,96 \cdot \frac{\sigma}{\sqrt{n}}
\end{aligned}$$

Sostituiamo con i valori di cui siamo a conoscenza e otteniamo:

$$\begin{aligned}
61 - 1,96 \cdot \frac{3}{\sqrt{5}} < \mu < 61 + 1,96 \cdot \frac{3}{\sqrt{5}} \\
58,37 < \mu < 63,63
\end{aligned}$$

Si ha che al 95% $\mu \in (58,37; 63,63)$.

Osservazione: all'aumentare di P aumenta l'ampiezza dell'intervallo $(\mu_1; \mu_2)$. Infatti considerando $p(\mu \in (\mu_1, \mu_2)) = 0,98$ si ha che:

$$p(-z \leq Z \leq z) = 0,98 \Rightarrow 2p(Z < z) - 1 = 0,98 \Rightarrow p(Z < z) = \frac{0,98 + 1}{2} = 0,9898$$

$z = 2,32$ (valore cercato nella tabella) $\bar{X} - 2,32 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 2,32 \cdot \frac{\sigma}{\sqrt{n}}$

$$\begin{aligned}
61 - 2,32 \cdot \frac{3}{\sqrt{5}} < \mu < 61 + 2,32 \cdot \frac{3}{\sqrt{5}} \\
57,89 < \mu < 64,11
\end{aligned}$$

Quindi $\mu \in (57,89; 64,11)$ al 98%.

Esercizi:

- 1) Un'impresa costruisce rasoi elettrici la cui durata si distribuisce normalmente. La durata media è di 6 anni e lo scarto quadratico medio è di 1,5 anni. Calcola:
 - a) La probabilità che un rasoio abbia una durata fino a 8 anni.
 - b) La probabilità che un rasoio abbia una durata di almeno 7 anni.
 - c) Il numero di rasoi che, su una produzione di 400 abbia una durata compresa tra i 4 e i 9 anni.

- 2) L'andamento dei prezzi di una crema solare sul mercato durante l'anno ha il comportamento di una variabile aleatoria di distribuzione normale, con media di 10 euro e scarto quadratico medio di 1,2 euro.
 - a) Determina l'intervallo monetario in cui si concentra il 95% dei casi del prezzo della crema.
 - b) Calcola la probabilità che il prezzo della crema sia compreso tra 9,5 e 10,5 euro.
 - c) Calcola la probabilità che il prezzo sia minore di 9 euro.

- 3) La fabbrica A produce matite colorate. Una prova su 100 matite scelte a caso ha indicato un peso medio di 12 grammi. Si supponga che il peso di una matita sia una variabile casuale distribuita secondo la legge normale, di parametri μ non nota e $\sigma^2 = 4$. Si determini l'intervallo di confidenza per μ pari al 95%.

