

V.a. continue

Ricordiamo:

DEFINIZIONE DI VARIABILE ALEATORIA

Una variabile aleatoria (in breve v.a.) X è una funzione che ha come dominio Ω e come codominio \mathbb{R} . In formule:

$$X : \Omega \rightarrow \mathbb{R}.$$

DEFINIZIONE DI V.A. DISCRETA O CONTINUA

Una v.a. $X : \Omega \rightarrow \mathbb{R}$ avente come immagine in \mathbb{R} l'insieme V (= insieme dei valori che X può assumere) si dice

- discreta se V è un insieme finito oppure infinito numerabile;
- continua se V è un insieme infinito continuo (= più numeroso del numerabile).

La densità di una v.a. continua

DEFINIZIONE DI DENSITÀ CONTINUA DI V.A.

Sia $X : \Omega \rightarrow \mathbb{R}$ una v.a. continua. Se esiste una funzione integrabile $f_X : \mathbb{R} \rightarrow [0, +\infty)$, tale che per ogni intervallo $J \subseteq \mathbb{R}$

$$\mathbb{P}(X \in J) = \int_J f_X(x) dx,$$

allora f_X è la funzione di densità (continua) della v.a. X e la X è detta v.a. assolutamente continua.

Le v.a. assolutamente continue spesso vengono dette continue, anche se esistono v.a. continue che non sono a.c. Ad esempio: scelgo una persona a caso e lancio una moneta. Se esce testa conto le dita della sua mano destra, se esce croce misuro l'altezza. Il risultato è una v.a. che ha come codominio una parte discreta (il numero di dita possibili) e una parte continua (le altezze possibili) ma non assolutamente continua (non esiste f_X con la proprietà della definizione).

Differenze con le discrete

Caso discreto

$f_X(a)$ è la probabilità che X assuma il valore a .

Caso continuo

$f_X(a)$ NON è la probabilità che X assuma il valore a : se X è continua allora

$$\mathbb{P}(X = a) = \int_a^a f_X(x) dx = 0.$$

Caso continuo

$$\mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b) - \mathbb{P}(X = a) = \int_a^b f_X(x) dx,$$

$$\mathbb{P}(a \leq X < b) = \mathbb{P}(a \leq X \leq b) - \mathbb{P}(X = b) = \int_a^b f_X(x) dx.$$

Differenze con le discrete

Caso discreto

$\mathbb{P}(X \in [a, b])$ è la **somma** delle $f_X(x)$ su tutti gli x valori possibili in $[a, b]$. In formule se X è discreta con insieme dei valori possibili V ,

$$\mathbb{P}(X \in [a, b]) = \sum_{x \in [a, b] \cap V} f_X(x).$$

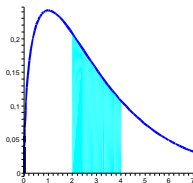
Caso continuo

$\mathbb{P}(X \in [a, b])$ è l'**integrale** di $f_X(x)$ su $[a, b]$. In formule (se X è continua),

$$\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x) dx.$$

Integrale come area

Può essere utile ricordare che $\int_a^b f(x)dx$ è l'area che sta “sopra” all'intervallo (a, b) e “sotto” la curva $y = f(x)$: ad es. qui $(a, b) = (2, 4)$



In particolare se $f_X(\cdot)$ è la densità di una v.a. continua, allora deve essere

$$\int_{\mathbb{R}} f_X(x)dx = \mathbb{P}(X \in \mathbb{R}) = 1.$$

Legge o distribuzione di una v.a.

Legge di una v.a.

Data una v.a. X , se per ogni intervallo reale I sappiamo calcolare

$$\mathbb{P}(X \in I)$$

diciamo che conosciamo la legge o distribuzione di X .

Conoscere la legge delle v.a. discrete e di quelle assolutamente continue equivale a conoscere la loro densità (come si calcola $\mathbb{P}(X \in I)$ nei due casi?).

A essere precisi la legge di una v.a. è una probabilità definita su \mathbb{R} .

V.a. identicamente distribuite

V.a. identicamente distribuite

Se le v.a. X_1, \dots, X_n hanno la stessa legge, si dice che sono identicamente distribuite o brevemente i.d.

In pratica significa che le v.a. in questione sono “dello stesso tipo” (ad esempio tutte $\mathcal{B}(0.5)$ oppure $\mathcal{N}(0, 1)\dots$).

Il valore atteso di una v.a. continua

DEFINIZIONE DI VALORE ATTESO DI X CONTINUA

Sia $X : \Omega \rightarrow \mathbb{R}$ una v.a. continua avente densità f_X . Se $\int_{\mathbb{R}} |x| \cdot f_X(x) dx < +\infty$ allora si definisce il valore atteso di X come il numero reale

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \cdot f_X(x) dx.$$

Ricordiamo e confrontiamo:

DEFINIZIONE DI VALORE ATTESO DI X DISCRETA

Sia $X : \Omega \rightarrow \mathbb{R}$ una v.a. discreta avente come immagine in \mathbb{R} l'insieme V . Se la serie $\sum_{v \in V} |v| \cdot f_X(v)$ converge, allora si chiama valore atteso di X il numero reale

$$\mathbb{E}(X) = \sum_{v \in V} v \cdot f_X(v).$$

La varianza

DEFINIZIONE DI VARIANZA DI UNA V.A.

Data una variabile aleatoria X la sua varianza è il numero reale:

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2).$$

Nota bene

La definizione di varianza NON cambia per v.a. discrete o continue: quello che cambia è COME SI CALCOLA il valore atteso \mathbb{E} .

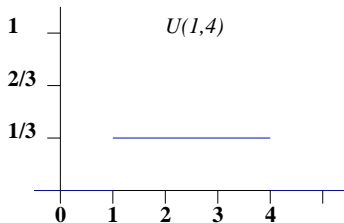
La v.a. uniforme

DEFINIZIONE DI V.A. $\mathcal{U}(a, b)$

Una v.a. uniforme sull'intervallo (a, b) è la v.a. continua che assume valori fra a e b e ha densità

$$f_X(x) = \frac{1}{b-a} \quad \text{se } x \in (a, b),$$

mentre $f_X(x) = 0$ se $x \notin (a, b)$.

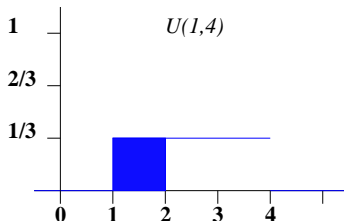


Un calcolo di probabilità

Se $X \sim \mathcal{U}(1, 4)$, qual è la probabilità che X sia minore di 2?

$$\mathbb{P}(X < 2) = \int_{-\infty}^2 f_X(x) dx.$$

Senza calcolare l'integrale, ricordiamo che si tratta dell'**area** sotto la funzione che integriamo:



La risposta è $1/3$.

Valore atteso della uniforme

Prendiamo $X \sim \mathcal{U}(1, 4)$:

$$\begin{aligned}\mathbb{E}(X) &= \int_1^4 x \cdot \frac{1}{3} dx \\ &= \left[\frac{x^2}{6} \right]_1^4 \\ &= \frac{16}{6} - \frac{1}{6} = \frac{5}{2} = 2.5.\end{aligned}$$

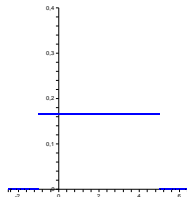
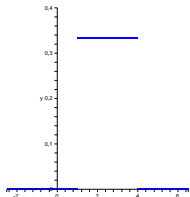
Quindi è il punto medio del segmento (1,4)!!! In generale

Valore atteso dell'uniforme

Se $X \sim \mathcal{U}(a, b)$, allora $\mathbb{E}(X) = \frac{a+b}{2}$.

Varianza della uniforme

Confrontiamo due uniformi:



quale ha varianza maggiore?

Risposta: quella sull'intervallo più largo.

Infatti

Varianza dell'uniforme

Se $X \sim \mathcal{U}(a, b)$, allora $\text{Var}(X) = \frac{(b-a)^2}{12}$.

Calcolo della varianza dell'uniforme

Calcolo

Se $X \sim \mathcal{U}(a, b)$, allora $(\mathbb{E}(X))^2 = \left(\frac{a+b}{2}\right)^2 = \frac{a^2+2ab+b^2}{2}$,

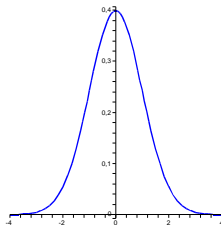
$$\begin{aligned}\mathbb{E}(X^2) &= \int_a^b x^2 \cdot \frac{1}{b-a} dx = \left[\frac{x^3}{3(b-a)} \right]_a^b \\ &= \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}.\end{aligned}$$

da cui

$$\begin{aligned}\text{Var}(X) &= \frac{b^2 + ab + a^2}{3} - \frac{a^2 + 2ab + b^2}{2} \\ &= \frac{4b^2 + 4ab + 4a^2 - 3a^2 - 6ab - 3b^2}{12} = \frac{(b-a)^2}{12}.\end{aligned}$$

La v.a. $\mathcal{N}(0, 1)$

La v.a. più importante della statistica è la normale (o gaussiana) standard: ecco il grafico della sua densità.



DEFINIZIONE DI V.A. $\mathcal{N}(0, 1)$

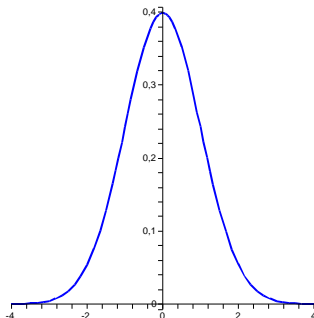
Una v.a. normale standard $\mathcal{N}(0, 1)$ è la v.a. continua che assume valori in tutto \mathbb{R} e ha densità come nel grafico sopra

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

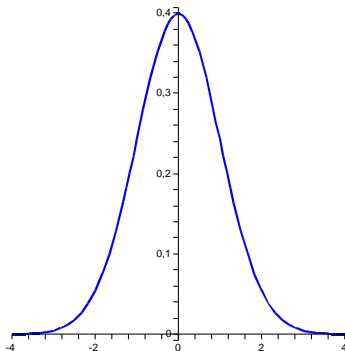
La v.a. $\mathcal{N}(0, 1)$

Cosa ricordare

Non importa l'espressione della densità! È invece **molto importante** conoscere il grafico e le sue proprietà.



Proprietà della densità di $\mathcal{N}(0, 1)$

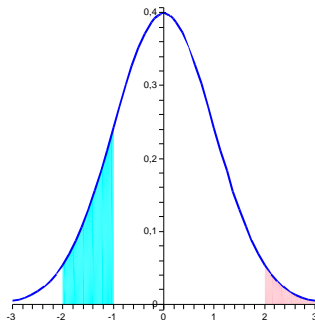


- Il massimo è in $x = 0$ (e $\max = \frac{1}{\sqrt{2\pi}} \approx 0.3989$).
- È simmetrica rispetto all'asse delle y .
- Va a zero a $\pm\infty$.
- $\mathbb{P}(|X| > 3) = 0.0027$.
- $\mathbb{P}(|X| > 4) = 0.000063$.
- $\mathbb{P}(-3 < X < 3) = 0.9973$.

Di fatto per le applicazioni si considera che la probabilità di assumere valori > 3 o < -3 sia trascurabile.

La probabilità

Dobbiamo pensare sempre alle probabilità come aree: ad esempio



in azzurro $\mathbb{P}(-2 \leq X \leq -1)$, in rosa $\mathbb{P}(X > 2)$.

La probabilità

Due notizie: una **buona** ed una **cattiva**.

La buona notizia è che queste aree non dobbiamo calcolarle (quindi niente integrale...) perché ci sono delle tavole (o anche pc che fanno il conto per noi...).

La cattiva notizia è che dobbiamo imparare a usare le tavole: esse riportano solo le $\mathbb{P}(\mathcal{N}(0, 1) \leq t)$, con $t \in [0, 2.99]$.

La funzione che a t associa $\mathbb{P}(\mathcal{N}(0, 1) \leq t)$ è indicata con Φ e si chiama **funzione di ripartizione della normale standard**.

Le tavole

La seguente tabella riporta i valori della funzione

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = P(Z \leq x) \text{ con } Z \sim \mathcal{N}(0, 1).$$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861

Esempio

La riga ci fa trovare il numero fino al primo decimale, la colonna il secondo decimale. Ecco ad esempio

$$\mathbb{P}(\mathcal{N}(0,1) \leq 1.53) = 0.93699.$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861

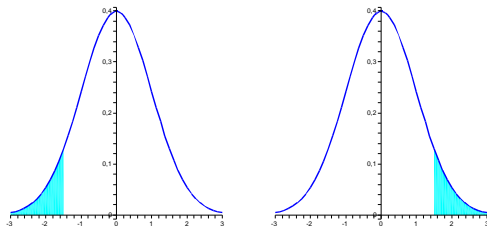
Le altre probabilità - 1

Si ottengono ragionando sulla simmetria del grafico.

Se $t > 0$

$$\mathbb{P}(\mathcal{N}(0, 1) \leq -t) = 1 - \mathbb{P}(\mathcal{N}(0, 1) \leq t) = 1 - \Phi(t).$$

Infatti l'area richiesta è quella colorata nel grafico di sinistra, che è uguale a quella colorata nel grafico di destra.

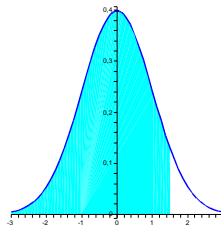
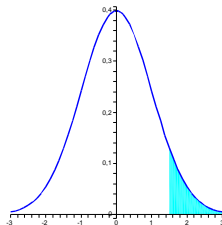


Le altre probabilità - 2

Per qualsiasi $t \in \mathbb{R}$

$$\mathbb{P}(\mathcal{N}(0, 1) > t) = 1 - \mathbb{P}(\mathcal{N}(0, 1) \leq t) = 1 - \Phi(t).$$

Infatti l'area richiesta è quella colorata nel grafico di sinistra, che è uguale a 1 - quella colorata nel grafico di destra.

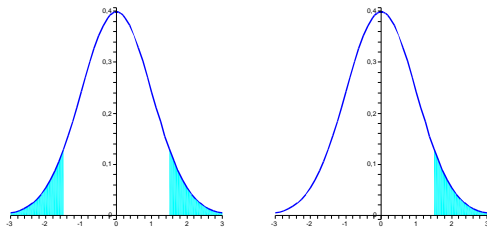


Le altre probabilità - 3

Se $t > 0$

$$\mathbb{P}(|\mathcal{N}(0, 1)| > t) = 2(1 - \mathbb{P}(\mathcal{N}(0, 1) \leq t)) = 2 - 2\Phi(t).$$

Infatti l'area richiesta è quella colorata nel grafico di sinistra, che è uguale a 2 volte quella colorata nel grafico di destra.

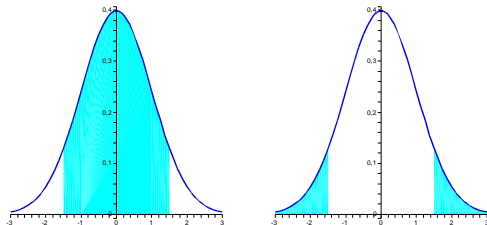


Le altre probabilità - 4

Se $t > 0$

$$\mathbb{P}(|\mathcal{N}(0, 1)| \leq t) = 1 - \mathbb{P}(|\mathcal{N}(0, 1)| > t) = 2\Phi(t) - 1.$$

Infatti l'area richiesta è quella colorata nel grafico di sinistra, che è uguale a 1 - quella colorata nel grafico di destra.



Le altre probabilità - 5

Se $a, b \in \mathbb{R}$, $-\infty < a \leq b < +\infty$ allora

$$\begin{aligned}\mathbb{P}(a < \mathcal{N}(0, 1) \leq b) &= \mathbb{P}(\mathcal{N}(0, 1) \leq b) - \mathbb{P}(\mathcal{N}(0, 1) \leq a) \\ &= \Phi(b) - \Phi(a).\end{aligned}$$

Valore atteso e varianza di $\mathcal{N}(0, 1)$

Teorema

Per $X \sim \mathcal{N}(0, 1)$ si ha $\mathbb{E}(X) = 0$ e $\text{Var}(X) = 1$.

Ricordiamo che $\mathbb{E}(X)$ sta ad indicare una media teorica (dove pesano di più i valori più probabili), mentre $\text{Var}(X)$ sta ad indicare la dispersione (pesata) dei valori.

Le v.a. normali qualsiasi

DEFINIZIONE DI V.A. $\mathcal{N}(\mu, \sigma^2)$

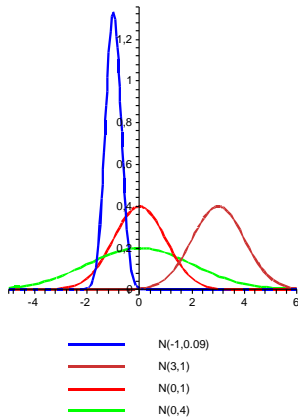
Sia $Z \sim \mathcal{N}(0, 1)$, e siano $\mu, \sigma \in \mathbb{R}$. La v.a. $X = \sigma Z + \mu$ si indica con $X \sim \mathcal{N}(\mu, \sigma^2)$ ed è detta v.a. normale di parametri μ e σ^2 .

Valore atteso e varianza $\mathcal{N}(\mu, \sigma^2)$

Sia $X \sim \mathcal{N}(\mu, \sigma^2)$. Allora $\mathbb{E}(X) = \mu$ e $\text{Var}(X) = \sigma^2$.

Nota: il fatto che valore atteso e varianza siano questi segue dalle proprietà di valore atteso e varianza.

Le diverse densità



Proprietà delle $\mathcal{N}(\mu, \sigma^2)$

La densità di $\mathcal{N}(\mu, \sigma^2)$ ha le proprietà:

- Il massimo è in $x = \mu$.
- L'altezza del massimo può anche superare 1 (dunque per le continue può essere $f_X(x) > 1$ per qualche x).
- Ha forma a campana.
- L'area sotto la curva vale 1.
- Se la varianza è minore di 1 la campana è stretta e alta.
- Se la varianza è maggiore di 1 la campana è larga e bassa.
- $\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$.

Quando si usa il modello normale

Esempi di situazioni in cui si usa:

- misure affette da errore (es. valore dell'accelerazione di gravità in un certo luogo)
- grandezze misurate sugli individui di una popolazione (es. statura, peso, lunghezza del braccio destro...)
- dimensioni di oggetti prodotti in serie con l'intento di produrli uguali.

Esempio di calcolo di probabilità

Il contenuto in ml delle bottiglie d'acqua di una certa marca sia $\mathcal{N}(1500, 25)$ (nel senso che, presa una bottiglia a caso, il suo contenuto è $X \sim \mathcal{N}(1500, 25)$).

Qual è la probabilità di acquistare una bottiglia il cui contenuto non supera 1480ml? e 1495ml? E che abbia contenuto compreso fra 1495ml e 1508ml?

Anzitutto scriviamo in formule le 3 domande:

$$\mathbb{P}(X \leq 1480) = ?$$

$$\mathbb{P}(X \leq 1495) = ?$$

$$\mathbb{P}(1495 \leq X \leq 1508) = ?$$

Esempio di calcolo

Abbiamo le tavole solo per $Z \sim \mathcal{N}(0, 1)$, quindi dobbiamo ricondurre le domande su X a domande su Z .

Ricordiamo: $X = \sigma Z + \mu$, dunque $Z = \frac{X - \mu}{\sigma}$.

Standardizzazione

Questa operazione (sottraggo il valore atteso e divido per la radice della varianza) è detta **standardizzazione** di X (infatti ottengo una variabile standard).

Quindi riscrivo le domande:

$$\mathbb{P}\left(\frac{X - 1500}{5} \leq \frac{1480 - 1500}{5}\right) = ?$$

$$\mathbb{P}\left(\frac{X - 1500}{5} \leq \frac{1495 - 1500}{5}\right) = ?$$

$$\mathbb{P}\left(\frac{1495 - 1500}{5} \leq \frac{X - 1500}{5} \leq \frac{1508 - 1500}{5}\right) = ?$$

Esempio di calcolo

$$\mathbb{P}\left(\frac{X - 1500}{5} \leq -4\right) = \Phi(-4) \approx 0$$

$$\begin{aligned}\mathbb{P}\left(\frac{X - 1500}{5} \leq -1\right) &= \Phi(-1) = 1 - \Phi(1) \\ &= 1 - 0.84134 = 0.15866\end{aligned}$$

$$\begin{aligned}\mathbb{P}\left(-1 \leq \frac{X - 1500}{5} \leq 1.6\right) &= \Phi(1.6) - \Phi(-1) \\ &= 0.94520 - 0.84134 \\ &= 0.10386\end{aligned}$$

Altro tipo di esercizi

L'esercizio appena visto è, in un certo senso, di tipo "diretto": si **conoscono i parametri** (valore atteso e varianza) della v.a. normale in questione, e si **calcola la probabilità** di un qualche evento.

Si incontrano frequentemente anche problemi di tipo "inverso": si **conosce la probabilità** di uno o più eventi e si **calcolano i parametri**.

Oppure si **conoscono i parametri**, si **fissa una probabilità** e si **determina un evento con quella probabilità**.

Esempio: calcoliamo la varianza

Supponiamo che il contenuto in ml delle bottiglie d'acqua di una certa marca sia $X \sim \mathcal{N}(1500, \sigma^2)$ (dunque conosciamo il valore atteso ma non la varianza). Si sappia inoltre che il 95% delle bottiglie contiene al massimo 1512ml. Quanto vale la varianza?

Il dato è

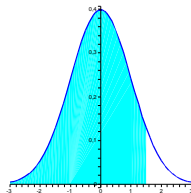
$$\mathbb{P}(X \leq 1512) = 0.95.$$

Le tavole

Standardizziamo (ricordiamo che $\frac{X-1500}{\sigma} \sim \mathcal{N}(0, 1)$):

$$\begin{aligned}\mathbb{P}\left(\frac{X - 1500}{\sigma} \leq \frac{1512 - 1500}{\sigma}\right) \\ = \mathbb{P}\left(\mathcal{N}(0, 1) \leq \frac{12}{\sigma}\right) = 0.95.\end{aligned}$$

Significa che se l'area turchese in figura vale 0.95, $12/\sigma$ deve essere il punto sull'asse delle x che divide la zona turchese da quella bianca.



Le tavole

Ricordiamo le tavole della funzione

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = P(Z \leq x) \text{ con } Z \sim \mathcal{N}(0, 1).$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861

Dovremo usare le tavole “all’inverso”: “dentro” cercheremo il valore più vicino possibile a 0.95, e guardando riga e colonna ricaveremo x.

I quantili

Una buona notizia: per le probabilità più usate questi numeri sono spesso presenti insieme alle tavole, ad esempio nel libro di testo o nell'eserciziario, sotto le tavole, troviamo la tabella dei quantili della legge normale standard:

α	0.90	0.95	0.975	0.99	0.995	0.999	0.9995
z_α	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905

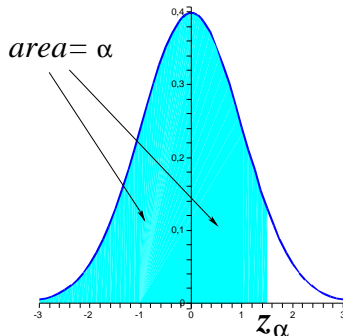
DEFINIZIONE DI QUANTILE DI $\mathcal{N}(0, 1)$

Sia $Z \sim \mathcal{N}(0, 1)$ e α un numero reale compreso fra 0 e 1. Si chiama α -quantile di Z il numero z_α tale che

$$\mathbb{P}(Z \leq z_\alpha) = \alpha.$$

I quantili

In pratica: si cerca il numero z_α alla cui sinistra l'area valga α . In figura: l'area turchese vale α , e z_α è il punto sull'asse delle x che divide la zona turchese da quella bianca.



Utilizzo nell'esercizio

Avevamo

$$\mathbb{P}\left(\mathcal{N}(0, 1) \leq \frac{12}{\sigma}\right) = 0.95.$$

Da cui si ricava

$$\frac{12}{\sigma} = z_{0.95} = 1.6449 \quad \implies \quad \sigma = \frac{12}{1.6449} = 7.2953.$$

La varianza è $\sigma^2 = 53.22$.

Esempio: cerchiamo l'evento

Con la definizione di quantile si possono risolvere diversi problemi (occorre vedere un po' di esercizi per farci un po' la mano). In tutti sfrutteremo la simmetria della campana della densità normale e il fatto che l'area totale vale 1.

Torniamo alle bottiglie di contenuto $X \sim \mathcal{N}(1500, 25)$. Il produttore vuole capire per quale intervallo di tipo $I = [1500 - \varepsilon, 1500 + \varepsilon]$ si abbia che almeno il 99% della produzione abbia contenuto compreso nell'intervallo I .

Esempio: cerchiamo l'evento

Si tratta di risolvere

$$\mathbb{P}(1500 - \varepsilon \leq X \leq 1500 + \varepsilon) \geq 0.99.$$

Standardizziamo:

$$\mathbb{P}\left(\frac{-\varepsilon}{5} \leq \frac{X - 1500}{5} \leq \frac{\varepsilon}{5}\right) \geq 0.99.$$

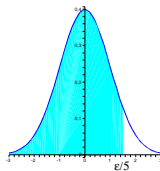
Scriviamo usando solo “aree a sinistra”:

$$\Phi\left(\frac{\varepsilon}{5}\right) - \Phi\left(\frac{-\varepsilon}{5}\right) \geq 0.99.$$

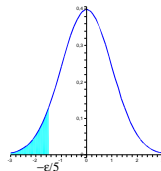
Esempio: cerchiamo l'evento

Il numero $\Phi\left(\frac{\varepsilon}{5}\right) - \Phi\left(\frac{-\varepsilon}{5}\right)$

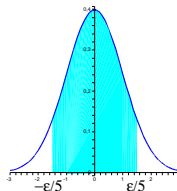
è l'area



meno l'area



quindi è l'area



Esempio: cerchiamo l'evento

Di queste aree sappiamo che

$$\Phi\left(\frac{\varepsilon}{5}\right) - \Phi\left(\frac{-\varepsilon}{5}\right) = 2\Phi\left(\frac{\varepsilon}{5}\right) - 1.$$

Quindi dobbiamo risolvere

$$2\Phi\left(\frac{\varepsilon}{5}\right) - 1 \geq 0.99$$

da cui

$$\Phi\left(\frac{\varepsilon}{5}\right) \geq 0.995,$$

ovvero

$$\frac{\varepsilon}{5} \geq z_{0.995} = 2.5758 \implies \varepsilon \geq 12.879.$$

L'intervallo più stretto possibile è $[1487.21, 1512.879]$.

Indipendenza di v.a. continue

DEFINIZIONE DI V.A. CONTINUE INDIPENDENTI

Date n v.a. continue X_1, X_2, \dots, X_n , esse si dicono indipendenti se vale l'uguaglianza

$$\begin{aligned}\mathbb{P}(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n) \\ = \mathbb{P}(X_1 \in I_1) \cdot \mathbb{P}(X_2 \in I_2) \cdots \mathbb{P}(X_n \in I_n),\end{aligned}$$

per ogni scelta di I_1, \dots, I_n intervalli reali.

Come per le discrete

Per dire che X_1, X_2, \dots, X_n sono indipendenti

non basta che valga

$$\begin{aligned}\mathbb{P}(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n) \\ = \mathbb{P}(X_1 \in I_1) \cdot \mathbb{P}(X_2 \in I_2) \cdots \mathbb{P}(X_n \in I_n),\end{aligned}$$

per qualche I_1, \dots, I_n , MA bisogna verificare l'uguaglianza
per ogni scelta di I_1, \dots, I_n .

Per dire che X_1, X_2, \dots, X_n NON sono indipendenti

basta trovare n intervalli I_1, \dots, I_n per cui

$$\begin{aligned}\mathbb{P}(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n) \\ \neq \mathbb{P}(X_1 \in I_1) \cdot \mathbb{P}(X_2 \in I_2) \cdots \mathbb{P}(X_n \in I_n).\end{aligned}$$

Indipendenza di v.a. continue

Valgono le proprietà viste per le indipendenti discrete:

Teorema

Se X e Y sono v.a. indipendenti, allora

- 1 il valore atteso del prodotto è il prodotto dei valori attesi: $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$;
- 2 la varianza della somma è la somma delle varianze: $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

In più per normali indipendenti si ha:

Se $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ e $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ sono v.a. indipendenti ed $a, b \in \mathbb{R}$, allora

- 3 $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$;
- 4 $aX + b \sim \mathcal{N}(a\mu_1 + b, a^2\sigma_1^2)$.