4 Standardizzazione

Abbiamo visto che, per risolvere il nostro problema guida relativo ai sondaggi (paragrafo 1.3), conviene ricondursi a calcolare probabilità relative alla distribuzione normale. Come fare? È quanto vedremo in questa sezione.

4.1 Standardizzazione della variabile aleatoria normale -video-

In questo video ti viene illustrato come esprimere probabilità relative ad una variabile aleatoria normale in termini di probabilità relative ad una variabile aleatoria speciale: la variabile aleatoria standard. Ti sono indicati anche i motivi sottesi ai vari passi, in modo che tu possa ricostruire il procedimento anche a distanza di tempo.

Il video si trova all'indirizzo https://youtu.be/M7Kt39446sU

4.2 Calcolo di probabilità relative alla distribuzione normale standard -video-

In questo video ti viene illustrato come calcolare probabilità relative alla variabile aleatoria normale standard, utilizzando opportune tavole. Questo video e il video 4.1, che ti consigliamo di esaminare, ti forniscono gli strumenti per calcolare probabilità relative alla distribuzione normale. Ti proponiamo anche una loro sintesi scritta nel paragrafo 4.3.

Il video si trova all'indirizzo https://youtu.be/Ywtt4tIZXVo

4.3 Calcolo di probabilità relative alla distribuzione normale

Tale procedimento è illustrato in dettaglio in due video (paragrafi 4.1 e 4.2) che ti consigliamo di esaminare. Ne proponiamo qui una sintesi.

Consideriamo una generica v.a. normale X di media μ e varianza σ^2. Dati due valori reali a, b con $a < b$, vogliamo calcolare

$$P(a \leq X \leq b)$$

Come visto nell'osservazione del paragrafo 3.1, tale probabilità non si può però calcolare mediante le formule "usuali".

Un modo per risolvere la questione è ricorrere a tavole che contengono valori (approssimati) di probabilità. Più precisamente essi sono relativi alla v.a. normale con media 0 e varianza 1, cioè alla v.a. standard Z. Per poter usare le tavole occorre dunque prima passare da valori di X a valori di Z.

Come fare? Proviamo a ragionare in termini di variabili aleatorie:

1. per avere *media*[a] 0, trasformiamo X nella nuova variabile $X - \mu$
2. per avere *varianza*[b] 1, trasformiamo $X - \mu$ nella nuova variabile $\frac{X-\mu}{\sigma}$

[a]Come la media della v.a. Z.
[b]Come la varianza della v.a. Z.

Pertanto, ogni v.a. normale X si può ricondurre[17] alla v.a. standard Z mediante la trasformazione

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

Per risolvere il problema iniziale consideriamo dunque i valori trasformati di a e b secondo tale relazione, ossia

$$a' = \frac{a - \mu}{\sigma} \quad \text{e} \quad b' = \frac{b - \mu}{\sigma}$$

Allora la probabilità richiesta si può così esprimere[a]:

$$P(a \leq X \leq b) = P(a' \leq Z \leq b')$$

[a]Tale risultato si può intuire per la costruzione fatta e si può comunque dimostrare formalmente (paragrafo 4.5).

[17]Nel paragrafo 4.4 discutiamo una dimostrazione formale dei due fatti enunciati. Ossia che la media e la varianza di $\frac{X-\mu}{\sigma}$ sono rispettivamente 0 e 1.

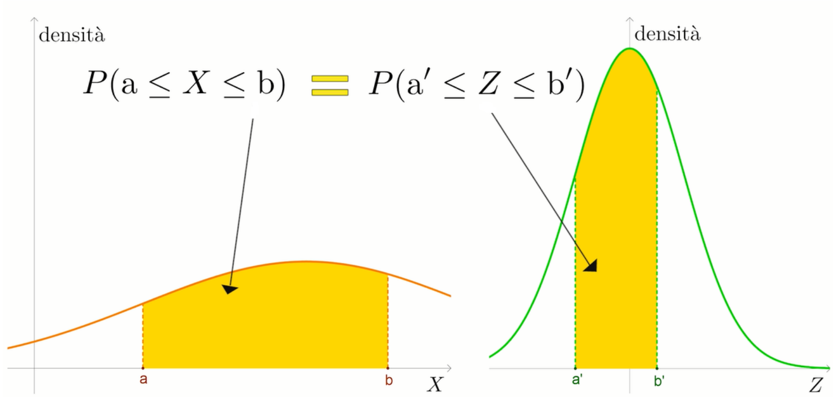

Figura 9: Vediamo qui un'interpretazione grafica, mediante le aree, dell'uguaglianza $P(a \le X \le b) = P(a' \le Z \le b')$.

Resta da determinare il valore di $P(a' \le Z \le b')$.

Possiamo ora utilizzare le tavole. Esse contengono valori approssimati di probabilità del tipo

$$P(Z \le k) \quad \text{dove } k \text{ è un valore positivo}$$

Per usare le tavole, si procede come suggerisce la figura 10 per l'esempio $P(Z \le 1,23)$.

Se invece occorre calcolare probabilità su altri intervalli, l'idea è di esprimere quanto richiesto in termini di probabilità del tipo[18] $P(Z \le k)$ con k positivo. Per far ciò, si possono utilizzare la simmetria della densità normale f e le proprietà degli integrali, interpretandole magari sul grafico di f in termini di *aree*.

Ad esempio, per calcolare $P(0,98 \le Z \le 1,74)$, che è una probabilità della forma in esame, si può procedere come illustrato in figura 11.

[18] Dato che quelle sono le probabilità che si trovano sulle tavole.

z	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

Figura 10: Lettura sulla tavola del valore approssimato di probabilità per $k = 1,23$. Quindi $P(Z \leq 1,23) \simeq 0,8907$.

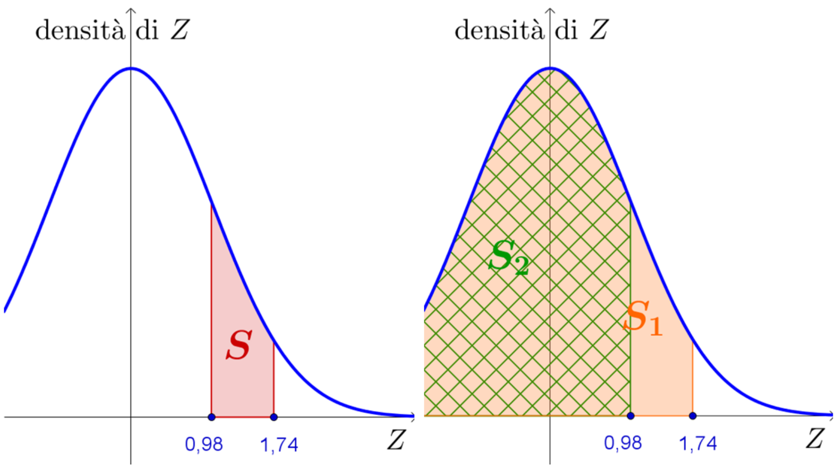

Figura 11: $\text{Area}(S) = \text{Area}(S_1) - \text{Area}(S_2) \implies P(0,98 \leq Z \leq 1,74) = P(Z \leq 1,74) - P(Z \leq 0,98)$.

4.4 Media e varianza della variabile aleatoria standardizzata -attività-

Sia X la variabile aleatoria normale di media μ e varianza σ^2.
Dimostra che la variabile aleatoria $\frac{X-\mu}{\sigma}$ ha media 0 e varianza 1.

Nota. Questo ci permette di concludere che $\frac{X-\mu}{\sigma}$ è[19] una variabile aleatoria con distribuzione normale standard.

Suggerimento

Usa le proprietà algebriche della media (varianza) per esprimere la media (varianza) di $\frac{X-\mu}{\sigma}$ in termini della media (varianza) della variabile X, dato che questa è nota. In particolare ricorda che per ogni numero c vale

$$\begin{array}{ll} E(cX) = cE(X) & V(cX) = c^2V(X) \\ E(X+c) = E(X) + c & V(c+X) = V(X) \end{array}$$

Con le notazioni $E(Y)$, $V(Y)$ indichiamo, come si usa spesso, rispettivamente la media e la varianza della v.a. Y.

[19] A rigore, per concludere che la variabile aleatoria così trasformata ha distribuzione normale standard, serve qualche considerazione aggiuntiva. Ma preferiamo non occuparcene.

Risoluzione

Partiamo col calcolare $E\left(\frac{X-\mu}{\sigma}\right)$. Sappiamo che $E(X) = \mu$, perciò cerchiamo di ricondurci alla variabile X, usando le proprietà algebriche della varianza:

$$E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E(X-\mu) = \frac{1}{\sigma}(E(X)-\mu) = \frac{1}{\sigma}(\mu-\mu) = 0$$

Abbiamo così provato che la media cercata è **0**.

Calcoliamo ora la $V\left(\frac{X-\mu}{\sigma}\right)$. Come prima cerchiamo di ricondurci alla variabile X visto che di questa sappiamo che la varianza è σ^2. Usando le proprietà algebriche della varianza si ha

$$V\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}V(X-\mu) = \frac{1}{\sigma^2}V(X) = \frac{1}{\sigma^2}\sigma^2 = 1$$

Abbiamo così provato che la varianza cercata è **1**.

4.5 Probabilità con la variabile aleatoria standardizzata -attività-

Sia X la variabile aleatoria normale di media μ e varianza σ^2. Dimostra che vale

$$P(a \leq X \leq b) = P(a' \leq Z \leq b')$$

dove Z è la variabile aleatoria normale standard e a' e b' sono i valori standardizzati[a] *di a e b.*

[a] Ossia $a' = \frac{a-\mu}{\sigma}$ e $b' = \frac{b-\mu}{\sigma}$.

Suggerimento

Scrivi esplicitamente le due probabilità $P(a \leq X \leq b)$, $P(a' \leq Z \leq b')$ come integrali delle densità delle rispettive variabili. Poi per provare l'uguaglianza richiesta, conviene utilizzare una opportuna sostituzione nell'integrale relativo alla v.a. X.

Risoluzione

Iniziamo esprimendo $P(a \leq X \leq b)$ come integrale della funzione densità di X, cioè

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, \mathrm{d}x$$

Analogamente

$$P(a' \leq Z \leq b') = \int_{a'}^{b'} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, \mathrm{d}z$$

Dobbiamo provare l'uguaglianza tra questi due integrali. Osservando gli esponenti nei due integrali oppure i loro estremi di integrazione, ci accorgiamo che conviene effettuare la sostituzione

$$z = \frac{x-\mu}{\sigma}$$

D'altronde si usa la stessa trasformazione per passare dalla v.a. X alla v.a. Z. Pertanto

$$x = \sigma z + \mu \quad \Longrightarrow \quad \mathrm{d}x = \sigma \mathrm{d}z$$

Con tale sostituzione si ha dunque

$$\begin{aligned} P(a \leq X \leq b) &= \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \, \mathrm{d}x \\ &= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2}} \, \sigma \mathrm{d}z \\ &= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, \mathrm{d}z \end{aligned}$$

Ma, come prima ricordato,

$$\int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, \mathrm{d}z = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$$

Abbiamo così provato formalmente che

$$P(a \leq X \leq b) = P(a' \leq Z \leq b')$$

dove

$$a' = \frac{a-\mu}{\sigma} \quad \text{e} \quad b' = \frac{b-\mu}{\sigma}$$