CAPITOLO QUINTO – DISTRIBUZIONE NORMALE

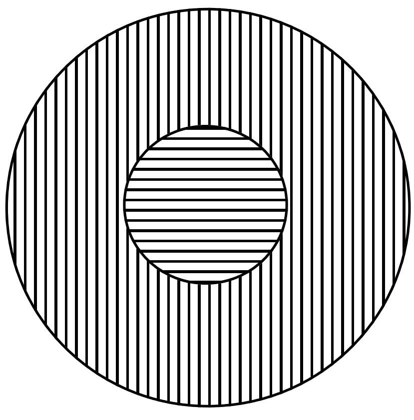
§1. Probabilità nel continuo

Fino ad ora abbiamo considerato casi in cui l’insieme degli eventi elementari è finito. Vediamo, mediante due semplici esempi, come si può procedere in alcune circostanze in cui questa condizione non è realizzata.

***Esempio* 1**. Marcello sa che il suo amico Carlo arriverà al bar Sport in un istante compreso fra le nove e le dieci di una data mattina. Egli decide di recarsi al bar alle nove e trenta e di attendere 10 minuti. Qual è la probabilità che Marcello incontri Carlo?
Per affrontare questo esempio, se si pone, come è spontaneo: e_t = “Carlo giunge al bar all’istante t” sorge una difficoltà: dato che vi sono infiniti istanti tra le nove e le dieci, gli eventi elementari sono in quantità infinita ed è alquanto problematico attribuire una probabilità a ciascuno di essi. Non si può porre $p(e_t) = 0$, in quanto vi è un istante in cui Carlo arriva; d’altra parte, se si pone $p(e_t) \neq 0$, anche molto piccolo, la somma delle probabilità degli infiniti eventi elementari sarebbe infinita (e non uguale a 1 come vorremmo). L’ostacolo può essere aggirato rinunciando ad attribuire una probabilità ai singoli eventi elementari e confrontando direttamente l’insieme dei casi favorevoli con l’insieme dei casi possibili. Si tratta di eseguire il rapporto fra l’intervallo tra le 9.30 e le 9.40 (l’insieme dei casi favorevoli all’incontro fra i due amici) e l’intervallo fra le 9 e le 10 (che costituisce l’insieme dei casi possibili).

Essendo evidentemente il primo intervallo pari a un sesto del totale, appare del tutto plausibile attribuire all’evento in questione probabilità $\frac{1}{6}$.

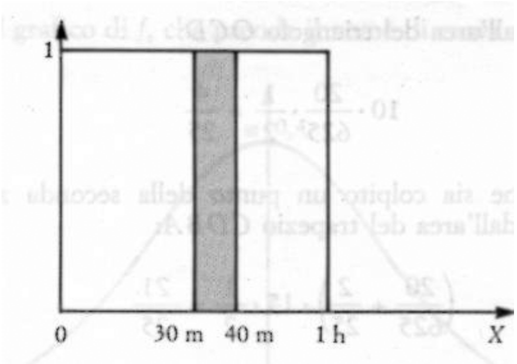
***Esempio* 2**. Un tiratore lancia una freccetta su un bersaglio circolare del raggio di 25 cm, il quale ha una zona centrale di raggio 10 cm. Se il tiratore colpisce il bersaglio, e tutti i punti di esso hanno la stessa probabilità di essere colpiti, qual è la probabilità che sia colpito un punto della zona centrale?
Come nell’esempio precedente, è assai difficile attribuire una probabilità all’evento “è colpito un punto del besaglio”. È possibile comunque confrontare l’area della zona centrale (insieme dei casi favorevoli) con l’area dell’intero bersaglio (insieme dei casi possibili).



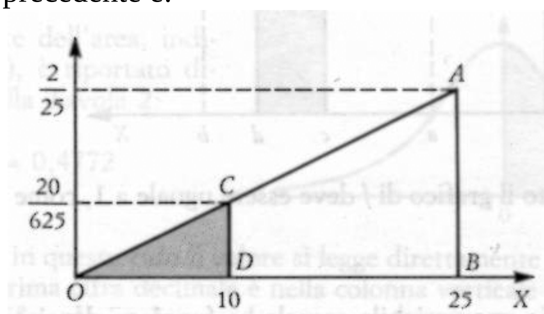
Posto E = “il tiratore colpisce la zona centrale”, si ha:

$$p(E) = \frac{\text{area zona centrale}}{\text{area bersaglio}} = \frac{\pi \cdot 10^2 \cdot \text{cm}^2}{\pi \cdot 25^2 \cdot \text{cm}^2} = \frac{4}{25}$$

Nel capitolo precedente ci siamo occupati di variabili casuali discrete (che assumono un numero finito di valori). In molti casi questa limitazione è troppo restrittiva in quanto la variabile casuale X può assumere tutti i valori di un intervallo di numeri reali.
Nell’Esempio 1 appare spontaneo introdurre una variabile casuale X, che assume tutti i valori compresi fra 0 e 1 ora e rappresenta l’istante in cui Carlo si troverà al bar Sport. Il problema che sorge con le variabili casuali continue è che *non è possibile attribuire una probabilità ai singoli valori di X* (come si è fatto per le variabili discrete), ma solo agli eventi del tipo “i valori di X sono compresi in un certo intervallo”. Procediamo come segue. Consideriamo il grafico seguente:

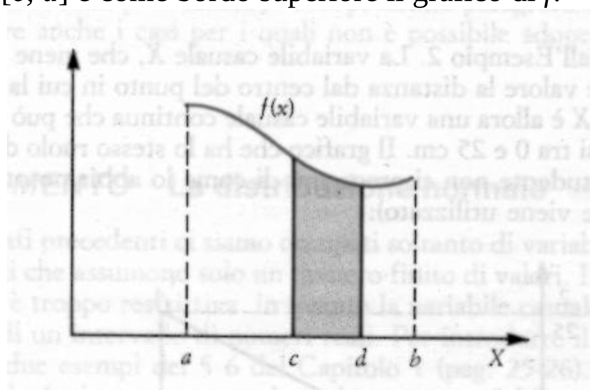


in cui in ascissa si riportano i valori di X e, in analogia con l’istogramma dell’Esempio 2 del Capitolo terzo relativo agli esisti dei lanci di un dado buono (essendo anche qui tutti gli istanti equiprobabili), tracciamo un segmento orizzontale in modo da ottenere un rettangolo (che in questo caso risulta un quadrato) di area 1.
La probabilità che Marcello incontri Carlo coincide con la probabilità che la variabile casuale X sia compresa fra 30 e 40 minuti: essa si può ottenere come l’area del rettangolo evidenziato in figura $\left(\text{ed è } \frac{1}{6}\right)$. Così la probabilità che Carlo arrivi nella prima mezz’ora, cioè che X sia compresa fra 0 e 30 minuti è $\frac{1}{2}$ ed è l’area del rettangolo avente come base tale intervallo. Osserviamo ancora che il lato superiore del rettangolo ha equazione: $f(x) = 1$ per $0 \le x \le 1$.
Passiamo ora all’Esempio 2. La variabile casuale X che viene spontaneo introdurre ha come valore la distanza dal centro del punto in cui la freccetta colpisce i bersaglio. X è allora una variabile casuale continua che può assumere tutti i valori compresi fra 0 e 25 cm. Il grafico che ha lo stesso ruolo del precedente è:



Per capire come lo abbiamo ottenuto si osservi in primo luogo che l'area del triangolo *OAB* è $25 \cdot \frac{2}{25} \cdot \frac{1}{2} = 1$. La probabilità che sia colpito un punto della zona centrale, ossia che $0 \le X \le 10$ è data dall'area del triangolo *OCD*: $10 \cdot \frac{20}{625} \cdot \frac{1}{2} = \frac{4}{25}$. La probabilità che sia colpito un punto della seconda zona (ossia che $10 \le X \le 25$), è data da $1 - \frac{4}{25} = \frac{21}{25}$ e corrisponde all'area del trapezio *CDBA* ossia $\left(\frac{60}{625} + \frac{2}{25}\right) \cdot 15 \cdot \frac{1}{2}$. Osserviamo infine che l'equazione del segmento *OA* è $f(x) = \frac{2x}{625}$ con $0 \le x \le 25$.

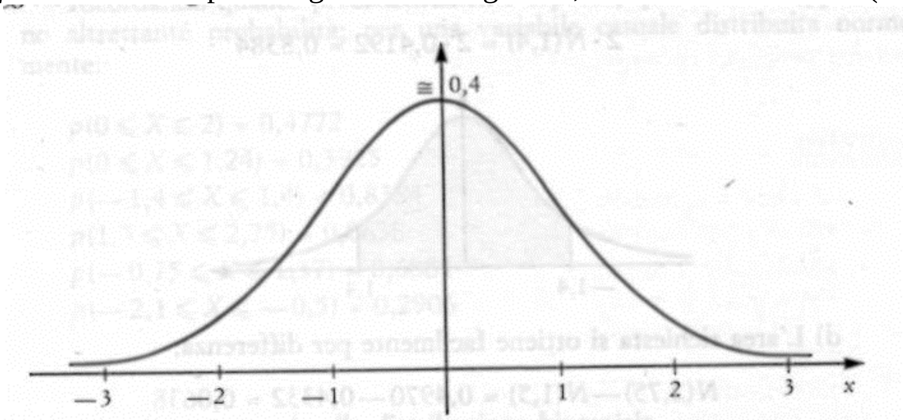
In generale, ad una variabile continua *X*, che può assumere tutti i valori tra *a* e *b*, anziché una distribuzione discreta di probabilità viene associata una funzione *f* a valori non negativi, detta **funzione densità di probabilità**, definita per $a \le x \le b$, tale che la probabilità che *X* assuma un valore compreso tra *c* e *d* (in formula $p(c \le X \le d)$) è uguale all'area della figura avente base l'intervallo $[c, d]$ e come bordo superiore il grafico di *f*:



La probabilità che *X* assuma un valore *k* qualsiasi dell'intervallo $[a, b]$ (in formula $p(X = k)$) è nulla (la superficie si riduce ad un segmento). L'area totale sotto il grafico di *f* deve essere uguale a 1, come nei due esempi precedenti.

§2. Distribuzione normale

Definizione. Se una variabile casuale ha densità di probabilità $f(x) = \frac{1}{2\pi} \cdot e^{-x^2/2}$ si dice che è **distribuita normalmente**. Non ha importanza in questa sede analizzare l'espressione analitica di *f*. Ciò che conta per il seguito è il suo grafico, detta **curva normale** (o **standard**):

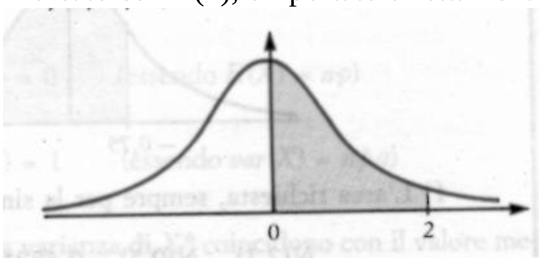


e, soprattutto, saper determinare le aree sotto tale grafico rispetto ad un qualsiasi intervallo. A tal fine sono predisposte delle tavole, una delle quali è riportata alla fine del capitolo. Nella tavola sono riportate le aree da 0 a z. Tuttavia, sfruttando la simmetria della curva rispetto all'asse y e ricordando che l'area totale sotto la curva normale è 1, si possono calcolare aree relativi a intervalli qualsiasi, come illustriamo nel seguente esempio.

***Esempio* 1**. Calcolare, servendosi della Tavola, le aree sottese dalla curva normale standard relative ai seguenti intervalli:

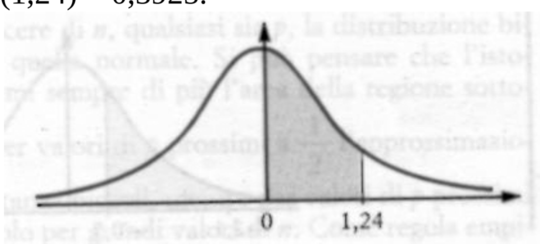
a)	[0; 2]	b)	[0; 1,24]	c)	[–1,4; 1,4]
d)	[1,5; 2,75]	e)	[–0,75; 1,37]	f)	[–2,1; –0,5]

a) Il valore dell'area, indicato con $N(2)$, è riportato direttamente sulla tavola:



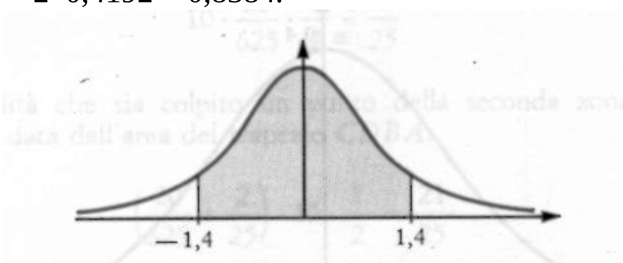
$N(2) = 0,4772$

b) Anche in questo caso il valore si legge direttamente sulla tavola (la prima cifra decimale è nella colonna verticale a sinistra, mentre la seconda cifra decimale va cercata nella prima riga orizzontale): $N(1,24) = 0,3925$.



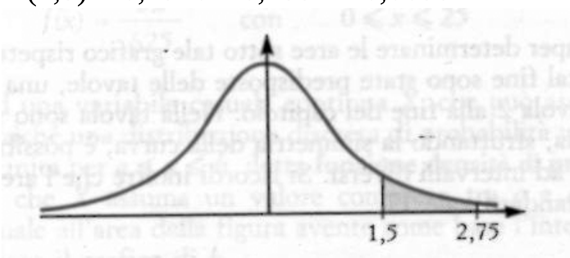
c) Data la simmetria della curva, l'area richiesta è evidentemente:

$2N(1,4) = 2\cdot 0,4192 = 0,8384$.



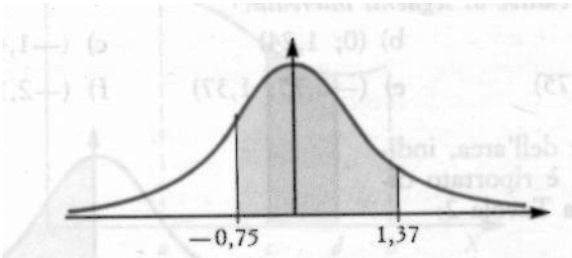
d) L'area richiesta si ottiene facilmante per differenza:

$N(2,75) – N(1,5) = 0,4970 – 0,4332 = 0,0638$



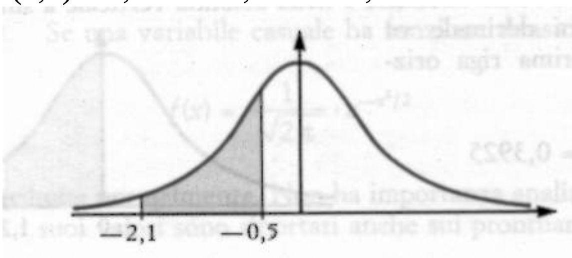
e) In questo caso occorre sommare l'area a destra, che vale $N(1,37)$, con l'area a sinistra che, per la simmetria della curva, vale $N(0,75)$. Si ha:

$N(1,37) + N(0,75) = 0,4147 + 0,2734 = 0,6881$



f) L’area richiesta, sempre per la simmetria della curva, vale:

$N(2{,}1) - N(0{,}5) = 0{,}4821 - 0{,}1915 = 0{,}2906$



Ricordando quanto precedentemente esposto, queste aree rappresentano altrettante probabilità; per la variabile casuale standard distribuita normalmente:

a) $p(0 \le X \le 2) = 0{,}4772$ b) $p(0 \le X \le 1{,}24) = 0{,}3925$
c) $p(-1{,}4 \le X \le 1{,}4) = 0{,}8384$ d) $p(1{,}5 \le X \le 2{,}75) = 0{,}0638$
e) $p(-0{,}75 \le X \le 1{,}37) = 0{,}6881$ f) $p(-2{,}1 \le X \le -0{,}5) = 0{,}2906$

§3. L’approssimazione normale alla distribuzione binomiale

La distribuzione normale (continua) è utile per approssimare le variabili casuali con distribuzione binomiale (discreta). La distribuzione normale ha valore medio 0 e varianza 1. Una variabile casuale X con distribuzione binomiale ha valore medio np e varianza npq. Introduciamo la nuova variabile casuale X^* così definita: $X^* = \frac{X - np}{\sqrt{npq}}$ detta ***standardizzata*** di X. Si può dimostrare che: $E(X^*) = \frac{E(X) - np}{\sqrt{npq}} = 0$ e $var(X^*) = \frac{1}{npq} \cdot var(X) = 1$, ossia che il valore medio e la varianza di X^* coincidono con il valore medio e la varianza della distribuzione normale. Un fondamentale teorema della teoria della probabilità (***Teorema di De Moivre-Laplace***) afferma che, al crescere di n, qualsiasi sia p, la distribuzione binomiale approssima sempre di più quella normale. Si può pensare che l’istogramma a rettangoli di X^* approssimi sempre meglio l’area della regione sottostante la curva normale standard. Per valori di p prossimi a $\frac{1}{2}$ l’approssimazione è già buona per valori di n abbastanza piccoli, mentre per valori di p prossimi a 0 o a 1 l’approssimazione è buona solo per grandi valori di n. Come regola empirica, l’approssimazione è soddisfacente se npq supera 10. Senza entrare in particolari, vedremo nei prossimi esempi come si può sfruttare questo risultato per risolvere problemi di prove ripetute.

***Esempio*1**. Si lancia una moneta 12 volte. Calcolare la probabilità di ottenere testa un numero di volte compreso fra 3 e 8 (estremi inclusi).
Risolviamo il problema dapprima mediante la tavola del capitolo precedente. Per $n = 12$, $p = 0{,}5$, la probabilità di ottenere esattamente tre volte testa è 0,0537; la probabilità di ottenerla 4 volte è 0,1208;...; di ottenerla 8 volte è 0,1208. In definitiva la probabilità richiesta è:

$0{,}0537 + 0{,}1208 + 0{,}1934 + 0{,}2256 + 0{,}1934 + 0{,}1208 = 0{,}9077$

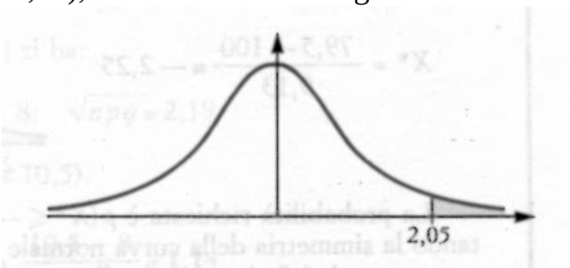
Per pervenire approssimativamente allo stesso risultato mediante la Tavola alla fine di questo capitolo si procede nel modo seguente. Come si è detto, la precedente somma equivale alla somma delle aree di tanti rettangoli che si estendono dal valore 2,5 al valore 8,5. Si calcolano allora i valori di X^* corrispondenti a 2,5 e 8,5 (in questo esempio $np = 6$ e $\sqrt{npq} = \sqrt{3} \cong$ 1,732):

per X = 2,5 si ha: $X = \dfrac{2{,}5 - 6}{1{,}732} \cong -2{,}02$

per X = 8,5 si ha: $X^* = \dfrac{8{,}5 - 6}{1{,}732} \cong 1{,}44$

La probabilità richiesta è l'area sotto la curva normale relativa all'intervallo di estremi –2,02 e 1,44: $p(-2{,}02 \leq X^* \leq 1{,}44) = N(1{,}44) + N(2{,}02) = 0{,}4251 + 0{,}4783 = 0{,}9034$.
L'approssimazione è buona pur essendo n = 12 un numero piccolo. Si ricordi comunque che per valori di p diversi da 1/2 l'approssimazione non sarebbe risultata così soddisfacente.

***Esempio* 2**. Si lancia una moneta 200 volte. Qual è la probabilità di ottenere un numero di teste superiore a 114?
Per questo esempio dobbiamo ricorerre alla Tavola in fondo al capitolo. Essendo $np = 100$ e $\sqrt{npq} \cong 7{,}07$, per X = 114,5 si ha: $X^* = \dfrac{114{,}5 - 100}{7{,}07} \cong 2{,}05$. La probabilità cercata si può approssimare con $p(X^* \geq 2{,}05)$, ossia con l'area in figura:



Se si osserva che metà dell'area sotto la curva normale è 0,5, l'area richiesta è $0{,}5 - N(2{,}05) = 0{,}5 - 0{,}4798 \cong 0{,}02$.

***Esempio* 3**. Si lancia 600 volte un dado. Qual è la probabilità di ottenere la faccia 6 un numero di volte:
a) compreso tra 90 e 110 (estremi inclusi);
b) inferiore a 80; c) superiore a 149?

La probabilità di successo è $p = 1/6$ (per cui $q = 5/6$), quindi $np = 100$ e $\sqrt{npq} \cong 9{,}13$.

a) Si deve determinare $p(89{,}5 \leq X \leq 110{,}5)$.

Per X = 89,5 si ha: $X^* = \dfrac{89{,}5 - 100}{9{,}13} \cong -1{,}15$

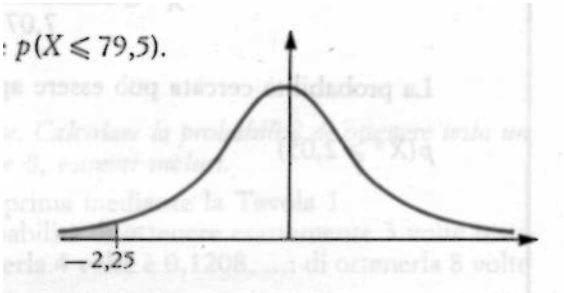
Per X = 110,5 si ha: $X^* = \dfrac{110{,}5 - 100}{9{,}13} \cong 1{,}15$

La probabilità $p(-1{,}15 \leq X^* \leq 1{,}15) = 2N(1{,}15) = 2 \cdot 0{,}3749 = 0{,}7498$.

b) Si deve determinare $p(X \leq 79{,}5)$.

Per X = 79,5 si ha: $X^* = \dfrac{79{,}5 - 100}{9{,}13} \cong -2{,}25$

La probabilità richiesta è $p(X^* \leq -2{,}25)$, ossia l'area in figura. Sfruttando la simmetria della curva normale e il fatto che metà dell'area ad essa sottostante è 0,5, si trova che l'area evidenziata è: $0{,}5 - N(2{,}25) = 0{,}5 - 0{,}4878 = 0{,}0122$.



c) Si deve determinare $p(X \geq 149{,}5)$.

Per X = 149,5 si ha: $X^* = \dfrac{149{,}5 - 100}{9{,}13} \cong 5{,}4$

Se si osserva che $p(X^* \geq 3{,}09) = 0{,}5 - 0{,}4990 = 0{,}001$, si deduce che è già bassissima la probabilità che X^* superi 3,09, per cui la probabilità che superi 5,4 si può ritenere praticamente nulla.
Questo risultato ha un'importante interpretazione. Supponiamo di lanciare 600 volte un dado e di ottenere la faccia 6 un numero di volte superiore a 149. Dato che, come si è visto, un siffatto evento ha una probabilità bassissima, possiamo affermare che, con elevatissima probabilità, il dado non è buono.

***Esempio* 4**. Il 40% della popolazione di una grande città è contrario ad un certo progetto del comune. Si sceglie un campione di 20 persone. Qual è la probabilità che nel campione vi siano almeno 11 persone contrarie al progetto?
Inizialmente osserviamo che, quando si sceglie un campione da una popolazione, si procede di solito con un'estrazione casuale senza reimmissione, ossia una persona non può essere scelta due volte. Ciò rende le successive estrazioni dipendenti. Tuttavia, in casi come questo, in cui la popolazione è elevata e il campione piccolo, si può procedere con ottima approssimazione considerando le successive estrazioni come un esperimento di Bernoulli.
Essendo $p = 0{,}4$ (e $q = 0{,}6$) si ha $np = 8$ e $\sqrt{npq} \cong 2{,}19$. Si vuole determinare $p(X \geq 10{,}5)$.

Per X = 10,5 si ha: $X^* = \dfrac{10{,}5 - 8}{2{,}19} \cong 1{,}14$

La probabilità richiesta è $p(X^* \geq 1{,}14) = 0{,}5 - N(1{,}14) = 0{,}5 - 0{,}3729 = 0{,}1271$.
Si può concludere che circa nel 13% dei casi il campione si esprimerebbe contro il progetto.

Negli esempi precedenti abbiamo sfruttato la distribuzione normale standard per approssimare valori di probabilità relativi a variabili casuali con distribuzione binomiale. In realtà, la distribuzione normale è in grado di approssimare molte altre importanti distribuzioni.
Molte variabili casuali, poi, hanno già una distribuzione normale con valore medio m e deviazione standard σ. Se si considera la variabile standardizzata $X^* = \dfrac{X - m}{\sigma}$ si può adoperare la Tavola alla fine del capitolo.

***Esempio* 5**. La vita media delle lampadine prodotte da una ditta è 1000 ore, con deviazione standard di 100 ore. Qual è la probabilità che, acquistando una lampadina di quella ditta, essa abbia una durata: a) superiore alle 1200 ore; b) inferiore alle 900 ore c) compresa fra le 950 e le 1150 ore?

Essendo $m = 1000$ e $\sigma = 100$, si ha: $X^* = \dfrac{X - 1000}{100}$.

a) Si chiede $p(X > 1200)$.

Per $X = 1200$ si ha: $X^* = \frac{1200 - 1000}{100} = 2$.

Quindi: $p(X > 1200) = p(X^* > 2) = 0{,}5 - N(2) = 0{,}5 - 0{,}4772 = 0{,}0228$.

b) Si chiede $p(X < 900)$.

Per $X = 900$ si ha: $X^* = \frac{900 - 1000}{100} = -1$

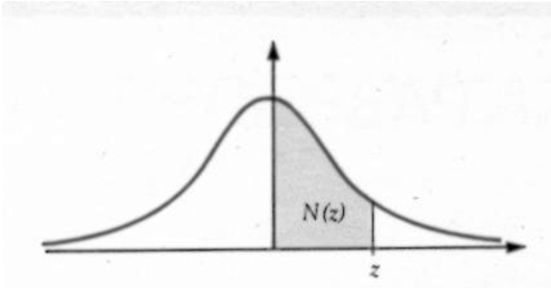
Quindi: $p(X < 900) = p(X^* < -1) = 0{,}5 - 0{,}3413 = 0{,}1587$.

c) Si chiede $p(950 < X < 1150)$.

Per $X = 950$ si ha: $X^* = \frac{950 - 1000}{100} = -0{,}5$

Per $X = 1150$ si ha: $X^* = \frac{1150 - 1000}{100} = 1{,}5$

Quindi: $p(950 < X < 1150) = p(-0{,}5 < X^* < 1{,}5) = N(1{,}5) + N(0{,}5) = 0{,}4332 + 0{,}1915 = 0{,}6247$.



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990