Azzolini Riccardo 2020-04-23

Variabili aleatorie continue

1 Problema: interventi urgenti

Problema: A un servizio di guardia medica arrivano in media, ogni ora, 3.5 richieste di eventi interventi urgenti a domicilio. Calcolare le probabilità:

1. che in una stessa ora arrivino 3, 4, 5 richieste urgenti;
2. che in una stessa ora arrivi un numero di richieste compreso tra 3 e 5;
3. che in una stessa ora arrivi un numero di richieste maggiore di 3;
4. che il tempo di arrivo della prima richiesta sia inferiore a 1 ora;
5. di attendere meno di 1 ora per avere 4 richieste.

Soluzioni:

Per rispondere alle prime tre domande, si definisce una variabile aleatoria discreta X che conta il numero di richieste urgenti ricevute nell'ora. È plausibile che X segua una distribuzione di Poisson, dato che sono verificate le condizioni del teorema che afferma quando essa si applichi:

- la probabilità di ricevere una richiesta in un intervallo di tempo piccolo è approssimativamente proporzionale all'ampiezza dell'intervallo;
- siccome le richieste urgenti sono eventi (relativamente) rari, la probabilità di riceverne due o più in un intervallo di tempo piccolo è molto inferiore rispetto a quella di riceverne una;
- ciascuna richiesta è indipendente dalle altre (scegliendo, per semplicità, di non contare eventuali richieste multiple relative a una stessa emergenza).

Il numero medio di richiesta che arrivano in un ora è $E(X) = 3.5$. La media della distribuzione Poisson(λ) coincide con il valore del parametro λ, quindi deve essere $\lambda = 3.5$, ovvero $X \sim \text{Poisson}(3.5)$.

Adesso, conoscendo la legge di X, si possono effettuare i calcoli:

1. Le probabilità di ricevere 3, 4 o 5 richieste urgenti nell'ora sono date direttamente dalla densità di X:

$$
\begin{aligned}
P\{X = k\} &= \frac{\lambda^k}{k!}e^{-\lambda} = \frac{3.5^k}{k!}e^{-3.5} \qquad k = 0, 1, 2, \dots \\
P\{X = 3\} &= \frac{3.5^3}{3!}e^{-3.5} \approx 0.2158 \\
P\{X = 4\} &= \frac{3.5^4}{4!}e^{-3.5} \approx 0.1888 \\
P\{X = 5\} &= \frac{3.5^5}{5!}e^{-3.5} \approx 0.1322
\end{aligned}
$$

2. La probabilità di ricevere, nell'ora, un numero di richieste compreso tra 3 e 5 è dato dalla somma delle tre probabilità appena calcolate:

$$
\begin{aligned}
P\{3 \le X \le 5\} &= P\{X = 3\} + P\{X = 4\} + P\{X = 5\} \\
&\approx 0.2158 + 0.1888 + 0.1322 = 0.5368
\end{aligned}
$$

3. Per trovare la probabilità di ricevere più di 3 richieste, si passa all'evento complementare:

$$
\begin{aligned}
P\{X > 3\} &= 1 - P\{X \le 3\} \\
&= 1 - (P\{X = 0\} + P\{X = 1\} + P\{X = 2\} + P\{X = 3\}) \\
&= 1 - \left(\frac{3.5^0}{0!}e^{-3.5} + \frac{3.5^1}{1!}e^{-3.5} + \frac{3.5^2}{2!}e^{-3.5} + \frac{3.5^3}{3!}e^{-3.5}\right) \\
&\approx 1 - (0.0302 + 0.1057 + 0.1850 + 0.2158) \\
&= 1 - 0.5367 \\
&= 0.4633
\end{aligned}
$$

   Osservazione: Quando servono tanti valori consecutivi della densità della Poisson, può essere utile la seguente relazione di ricorrenza, che permette di "riutilizzare" ciascun risultato nel calcolo del successivo:

$$
\begin{aligned}
P\{X = 0\} &= e^{-\lambda} \\
P\{X = k + 1\} &= \frac{\lambda}{k + 1}P\{X = k\}
\end{aligned}
$$

Invece, per le ultime due domande, si può ragionare direttamente sui tempi di attesa, nel continuo, sfruttando il legame tra Poisson ed esponenziale, oppure si possono trovare degli eventi discreti equivalenti. A scopo illustrativo, in seguito vengono discusse entrambe le soluzioni.

4. Sia Y la variabile aleatoria continua che esprime il tempo di arrivo della prima richiesta (misurato in ore). L'evento considerato è

$$\{Y < 1\} = \text{“il tempo di arrivo della prima richiesta è inferiore a 1 ora”}$$

Si osserva, però, che tale evento si verifica se e solo se arriva almeno una richiesta nell'ora, quindi esso è equivalente all'evento $\{X \geq 1\}$, e allora:

$$\begin{aligned} P\{Y \leq 1\} &= P\{X \geq 1\} \\ &= 1 - P\{X < 1\} \\ &= 1 - P\{X = 0\} \\ &= 1 - \frac{\lambda^0}{0!} e^{-\lambda} \\ &= 1 - e^{-\lambda} \\ &= 1 - e^{-3.5} \approx 1 - 0.0302 = 0.9698 \end{aligned}$$

In questo calcolo, è emersa la formula

$$P\{Y \leq 1\} = 1 - e^{-\lambda}$$

che corrisponde alla funzione di ripartizione della distribuzione esponenziale di parametro λ,

$$P\{Y \leq y\} = 1 - e^{-\lambda y}$$

valutata in $y = 1$.

Infatti, come visto in precedenza, si ha in generale che, quando i tempi di attesa tra eventi successivi sono esponenziali di parametro λ, il numero di eventi in un intervallo di tempo di ampiezza T è contato da una Poisson di parametro λT. Viceversa, partendo da una Poisson di parametro λ, i tempi di attesa sono esponenziali di parametro $\frac{\lambda}{T}$. In questo problema, $T = 1$ (sempre misurando in ore), dunque Y è esponenziale di parametro $\frac{3.5}{1} = 3.5$: $Y \sim \text{Exp}(3.5)$.

5. Siano Y_1, Y_2, Y_3, Y_4 i tempi di arrivo di ciascuna delle 4 richieste (misurati a partire dalla richiesta precedente). L'evento considerato è $\{Y_1 + Y_2 + Y_3 + Y_4 < 1\}$.

Ragionando direttamente nel continuo, si ha che $Y_i \sim \text{Exp}(3.5)$ (come per la domanda precedente), e perciò la loro somma segue una legge gamma:

$$Y_1 + Y_2 + Y_3 + Y_4 \sim \Gamma(4, 3.5)$$

Essendo questa una gamma con $\alpha = 4$ intero positivo, è possibile scrivere l'espressione della sua funzione di ripartizione,

$$F_4(y) = 1 - \sum_{k=0}^{3} \frac{(\lambda y)^k}{k!} e^{-\lambda y} = 1 - \sum_{k=0}^{3} \frac{(3.5y)^k}{k!} e^{-3.5y}$$

e usarla per calcolare la probabilità di ricevere 4 richieste in meno di un'ora:

$$
\begin{aligned}
P\{Y_1 + Y_2 + Y_3 + Y_4 < 1\} = F_4(1) &= 1 - \sum_{k=0}^{3} \frac{3.5^k}{k!} e^{-3.5} \\
&= 1 - \left(\frac{3.5^0}{0!} e^{-3.5} + \frac{3.5^1}{1!} e^{-3.5} + \frac{3.5^2}{2!} e^{-3.5} + \frac{3.5^3}{3!} e^{-3.5} \right) \\
&\approx 0.4633
\end{aligned}
$$

Si osserva che questo è esattamente il calcolo effettuato per $P\{X > 3\}$. Infatti, ricevere 4 richieste in meno di un'ora equivale a ricevere almeno 4 richieste nell'intera ora, cioè all'evento $\{X \geq 4\} = \{X > 3\}$, la cui probabilità, usando l'evento complementare e la funzione di ripartizione della Poisson,

$$
F_X(x) = \sum_{k=0}^{x} \frac{\lambda^k}{k!} e^{-\lambda}
$$

si calcola appunto come:

$$
\begin{aligned}
P\{X > 3\} &= 1 - P\{X \leq 3\} \\
&= 1 - F_X(3) \\
&= 1 - \sum_{k=0}^{3} \frac{3.5^k}{k!} e^{-3.5} \\
&= F_4(1) = P\{Y_1 + Y_2 + Y_3 + Y_4 < 1\}
\end{aligned}
$$

2 Problema: assicurazione

Problema: Una compagnia di assicurazioni ha 3840 assicurati. Per ciascuno di essi, la probabilità che denunci almeno un incidente all'anno è $p = \frac{1}{1200}$.

1. Trovare la probabilità che 0, 1, 2, 3, 4, ecc. assicurati denuncino almeno un incidente all'anno.
2. Qual è la probabilità che la prima denuncia arrivi entro l'anno?
3. Qual è il tempo che intercorre, in media, tra due denunce di assicurati diversi?

Soluzioni:

Per prima cosa, è importante notare che non si vogliono contare le singole denunce, bensì gli assicurati diversi che denunciano almeno un incidente (ma, potenzialmente, anche più di uno) nel corso dell'anno. Quindi, ciascun assicurato o si conta (se denuncia almeno un incidente) o no. Sia X la variabile aleatoria discreta che esprime il valore di tale conteggio.

Questa situazione è uno schema successo-insuccesso, nel quale ogni "prova" rappresenta un assicurato e ha "successo" se l'assicurato denuncia almeno un incidente nell'anno. Perciò, X segue una distribuzione binomiale di parametri $n = 3840$ e $p = \frac{1}{1200}$:

$$X \sim B\left(3840, \frac{1}{1200}\right)$$

Tuttavia, per effettuare calcoli su questa distribuzione sarebbe necessario gestire fattoriali molto grandi e valori di p^k molto piccoli. Invece, c'è un'alternativa più pratica: siccome n è piuttosto grande, e il valore di p è molto piccolo rispetto a n, la binomiale è ben approssimata dalla Poisson di parametro

$$\lambda = np = \frac{3840}{1200} = \frac{16}{5} = 3.2$$

quindi è ragionevole porre $X \sim \text{Poisson}(3.2)$.

Adesso, si può procedere con i calcoli:

1. La probabilità che esattamente $k = 0, 1, \dots$ assicurati denuncino almeno un incidente nell'anno è data dalla densità della Poisson:

   $$P\{X = k\} = \frac{3.2^k}{k!} e^{-3.2}$$

   Volendo, si potrebbe calcolare ciascun valore della densità separatamente, oppure usare la relazione di ricorrenza data prima, ma esiste anche un altro modo: siccome la distribuzione di Poisson ha molte importanti applicazioni, sono disponibili delle tavole numeriche che forniscono i valori della sua funzione di ripartizione per diversi λ, e, come già visto, dalla funzione di ripartizione è possibile ricavare anche la densità:

   $$P\{X = k\} = F_X(k) - F_X(k - 1)$$

   Ad esempio, nella tavola, sulla riga relativa a $\lambda = 3.2$ si legge che:

   $$\begin{array}{lll} F_X(0) \approx 0.0408 & F_X(1) \approx 0.1712 & F_X(2) \approx 0.3799 \\ F_X(3) \approx 0.6025 & F_X(4) \approx 0.7806 & \dots \end{array}$$

   Allora, si possono calcolare:

   $$\begin{aligned} P\{X = 0\} &= F_X(0) \approx 0.0408 \\ P\{X = 1\} &= F_X(1) - F_X(0) \approx 0.1712 - 0.0408 = 0.1304 \\ P\{X = 2\} &= F_X(2) - F_X(1) \approx 0.3799 - 0.1712 = 0.2087 \\ P\{X = 3\} &= F_X(3) - F_X(2) \approx 0.6025 - 0.3799 = 0.2226 \\ P\{X = 4\} &= F_X(4) - F_X(3) \approx 0.7806 - 0.6025 = 0.1781 \\ &\vdots \end{aligned}$$

2. La prima denuncia arriva entro l'anno se e solo se nell'anno ci sono complessivamente uno o più assicurati che denunciano:

$$\begin{aligned} P\{X \geq 1\} &= 1 - P\{X < 1\} \\ &= 1 - P\{X = 0\} \\ &= 1 - e^{-3.2} \\ &\approx 1 - 0.0408 \\ &= 0.9592 \end{aligned}$$

3. Per calcolare il tempo di attesa medio tra una denuncia e la successiva (sempre di assicurati diversi, cioè senza considerare eventuali denunce multiple fatte da uno stesso assicurato), è necessario definire una variabile aleatoria continua Y che misuri tale tempo di attesa, e calcolarne il valore medio $E(Y)$.

   Per il legame tra la Poisson e l'esponenziale, siccome si considera un intervallo di tempo unitario (1 anno), si ha che $Y \sim \text{Exp}(3.2)$. Ci sono vari modi calcolare il valore medio; il più semplice è usare direttamente la definizione:

$$\begin{aligned} E(Y) &= \int_{-\infty}^{+\infty} x f(x)\, dx \\ &= \int_{0}^{+\infty} x \cdot \lambda e^{-\lambda x}\, dx \\ &= \left[-x e^{-\lambda x}\right]_0^{+\infty} - \int_0^{+\infty} (-e^{-\lambda x})\, dx \\ &= \lim_{b \to +\infty} \left[-x e^{-\lambda x}\right]_0^b + \left[\frac{e^{-\lambda x}}{-\lambda}\right]_0^{+\infty} \\ &= \lim_{b \to +\infty} (\underbrace{-b e^{-\lambda b}}_{\to 0} + 0) + \lim_{b \to +\infty} \left[\frac{e^{-\lambda x}}{-\lambda}\right]_0^b \\ &= 0 + \lim_{b \to +\infty} \left(\frac{e^{-\lambda b}}{-\lambda} - \frac{1}{-\lambda}\right) \\ &= \lim_{b \to +\infty} \frac{1}{\lambda}(1 - \underbrace{e^{-\lambda b}}_{\to 0}) \\ &= \frac{1}{\lambda} \end{aligned}$$

$$\begin{gathered} g'(x) = \lambda e^{-\lambda x} \implies g(x) = -e^{-\lambda x} \\ h(x) = x \implies h'(x) = 1 \end{gathered}$$

   Allora, in questo caso, il tempo di attesa medio è:

$$E(Y) = \frac{1}{\lambda} = \frac{1}{3.2} = 0.3125$$

3 Problema: errore in un esperimento

Problema: In certi esperimenti, l'errore commesso nella determinazione della solubilità di una sostanza è una variabile aleatoria X avente distribuzione uniforme, con $a = -0.025$ e $b = 0.025$. Trovare la probabilità che l'errore:

1. sia compreso tra 0.010 e 0.015;
2. sia compreso tra -0.012 e 0.012.

Soluzioni:

La funzione di ripartizione di X,

$$F_X(x) = \begin{cases} 0 & \text{se } x \leq -0.025 \\ \dfrac{x+0.025}{0.025+0.025} = \dfrac{x+0.025}{0.05} = 20x + 0.5 & \text{se } -0.025 < x < 0.025 \\ 1 & \text{se } x \geq 0.025 \end{cases}$$

può essere usata per calcolare entrambe le probabilità richieste:

$$\begin{aligned} P\{0.010 < X < 0.015\} &= F_X(0.015) - F_X(0.010) \\ &= 20 \cdot 0.015 + 0.5 - 20 \cdot 0.010 - 0.5 \\ &= 0.3 - 0.2 \\ &= 0.1 \\ P\{-0.012 < X < 0.012\} &= F_X(0.012) - F_X(-0.012) \\ &= 20 \cdot 0.012 + 0.5 - 20(-0.012) - 0.5 \\ &= 0.24 + 0.24 \\ &= 0.48 \end{aligned}$$

4 Problema: parametri della distribuzione uniforme

Problema: La variabile aleatoria X è distribuita uniformemente nell'intervallo (a, b). Sapendo che

$$P\{X < 3\} = \frac{1}{4} \qquad P\{X < 7\} = \frac{3}{4}$$

calcolare a e b.

Soluzione:

Siccome X segue una distribuzione uniforme, la sua funzione di ripartizione deve essere della forma:

$$F_X(x) = \begin{cases} 0 & \text{se } x \leq a \\ \dfrac{x-a}{b-a} & \text{se } a < x < b \\ 1 & \text{se } x \geq b \end{cases}$$

Al momento, sono noti due suoi valori:

$$\begin{cases} F_X(3) = P\{X < 3\} = \dfrac{1}{4} \\ F_X(7) = P\{X < 7\} = \dfrac{3}{4} \end{cases}$$

Siccome questi valori non sono né 0 né 1, si deduce che i punti 3 e 7 appartengono all'intervallo (a, b). Allora, a e b possono essere determinati risolvendo il seguente sistema di equazioni:

$$\begin{cases} \dfrac{3-a}{b-a} = \dfrac{1}{4} \\ \dfrac{7-a}{b-a} = \dfrac{3}{4} \end{cases}$$

$$\begin{cases} 3 - a = \dfrac{1}{4}(b-a) \\ 7 - a = \dfrac{3}{4}(b-a) \end{cases} \qquad \text{(sicuramente } b - a \neq 0\text{: l'intervallo contiene almeno 3 e 7)}$$

$$\begin{cases} 4 \cdot 3 - 4a = b - a \\ 4 \cdot 7 - 4a = 3b - 3a \end{cases}$$

$$\begin{cases} 12 - 4a + a = b \\ 28 = 3b + 4a - 3a \end{cases}$$

$$\begin{cases} b = 12 - 3a \\ 28 = 3(12 - 3a) + a \end{cases}$$

$$\begin{cases} b = 12 - 3a \\ 28 = 36 - 9a + a \end{cases}$$

$$\begin{cases} b = 12 - 3a \\ 8a = 36 - 28 \end{cases}$$

$$\begin{cases} b = 12 - 3a \\ a = \dfrac{8}{8} = 1 \end{cases}$$

$$\begin{cases} b = 12 - 3 = 9 \\ a = 1 \end{cases}$$

Quindi, X è uniformemente distribuita in $(1, 9)$, ovvero ha la funzione di ripartizione

$$F_X(x) = \begin{cases} 0 & \text{se } x \leq 1 \\ \dfrac{x-1}{8} & \text{se } 1 < x < 9 \\ 1 & \text{se } x \geq 9 \end{cases}$$

e la densità

$$f(x) = \begin{cases} \dfrac{1}{8} & \text{se } 1 < x < 9 \\ 0 & \text{altrimenti} \end{cases}$$

5 Distribuzione normale

Definizione: La densità di probabilità **normale** o **gaussiana** di parametri μ e σ, con $\sigma > 0$, è:

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

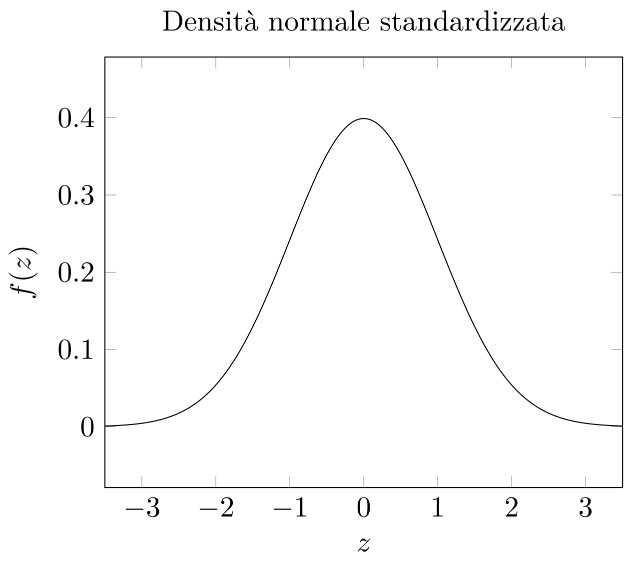
μ è il valore medio della distribuzione, e σ è la deviazione standard (scarto quadratico medio), quindi la varianza è σ^2.

Definizione: La densità di probabilità della **variabile normale standardizzata** Z è data da

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

ovvero è la normale con media $\mu = 0$ e deviazione standard $\sigma = 1$.

Il grafico della densità di una distribuzione normale ha una tipica forma a "campana" (centrata sulla media μ, e tanto più larga quanto più è ampia la deviazione standard σ):



La funzione di ripartizione $F_Z(z)$ di Z (e, in generale, quella di una distribuzione normale) non ha un'espressione analitica, quindi si usano tipicamente delle tavole che ne riportano i valori numerici. L'uso di queste tavole richiede però alcuni accorgimenti:

- Nelle tavole sono presenti solo i valori di $F_Z(z)$ per $z \geq 0$: quelli per $z < 0$ possono essere ricavati sfruttando la simmetria della densità normale standardizzata.
- Se si sta considerando una variabile normale X con parametri $\mu \neq 0$ e/o $\sigma \neq 1$, bisogna prima passare alla variabile standardizzata, nel modo seguente:

$$Z = \frac{X - \mu}{\sigma}$$
$$P\{a < X < b\} = P\left\{\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right\} = P\left\{\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right\}$$

6 Problema: diametro delle sfere

Problema: Il diametro effettivo delle sfere di acciaio prodotte da una ditta può essere considerato una variabile aleatoria normale X di media $\mu = 5.1\,\text{cm}$ e deviazione standard $\sigma = 0.1\,\text{cm}$.

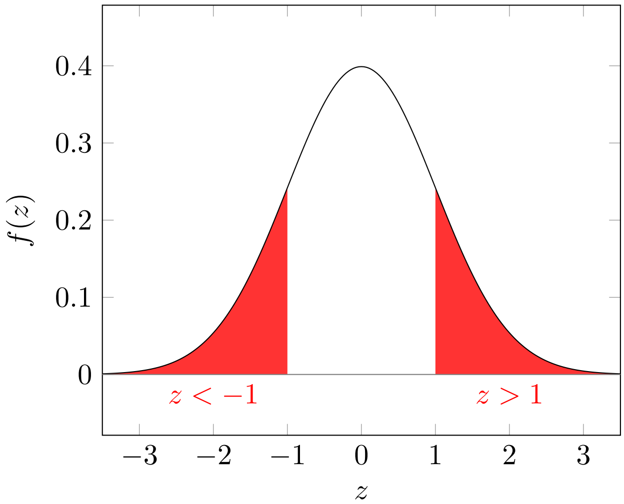
1. Calcolare la probabilità che il diametro di una sfera scelta a caso sia compreso tra 5.0 e 5.2 cm.
2. Calcolare la stessa probabilità, supponendo però che la deviazione standard sia $\sigma = 0.5\,\text{cm}$.

Soluzioni:

1. L'evento richiesto è $\{5.0 < X < 5.2\}$. Per calcolarne la probabilità, si passa alla variabile standardizzata Z:

$$\begin{aligned} P\{5.0 < X < 5.2\} &= P\left\{\frac{5.0 - 5.1}{0.1} < Z < \frac{5.2 - 5.1}{0.1}\right\} \\ &= P\{-1 < Z < 1\} \\ &= F_Z(1) - F_Z(-1) \end{aligned}$$

Il valore di $F_Z(1)$ è reperibile direttamente nelle tavole, ed è $F_Z(1) \approx 0.8413$. Invece, per ricavare $F_Z(-1)$ bisogna sfruttare, come anticipato, la simmetria[1] della densità. Ricordando che la funzione di ripartizione (e, più in generale, la probabilità di un intervallo), in quanto integrale della densità, corrisponde all'area sottostante al grafico di quest'ultima, si osserva che l'area per $z < -1$ è uguale a quella per $z > 1$,



e quindi:

$$\begin{aligned} F_Z(-1) &= P\{Z < -1\} \\ &= P\{Z > 1\} && \text{(simmetria)} \\ &= 1 - P\{Z < 1\} && \text{(evento complementare)} \\ &= 1 - F_Z(1) \\ &\approx 1 - 0.8413 \\ &= 0.1587 \end{aligned}$$

Adesso, si hanno tutti i dati necessari per calcolare la probabilità richiesta:

$$P\{5.0 < X < 5.2\} = F_Z(1) - F_Z(-1) \approx 0.8413 - 0.1587 = 0.6826$$

[1] Quello illustrato in seguito è solo uno dei possibili ragionamenti sulla simmetria della densità (che, comunque, portano tutti allo stesso risultato).

2. Nel secondo caso, siccome cambia solo il parametro $\sigma = 0.5\,\mathrm{cm}$, i calcoli sono analoghi:

$$\begin{aligned}
P\{5.0 < X < 5.2\} &= P\left\{\frac{5.0 - 5.1}{0.5} < Z < \frac{5.2 - 5.1}{0.5}\right\} \\
&= P\{-0.2 < Z < 0.2\} \\
&= F_Z(0.2) - F_Z(-0.2) \\
&= F_Z(0.2) - (1 - F_Z(0.2)) && \text{(simmetria)} \\
&= 2F_Z(0.2) - 1 \\
&\approx 2 \cdot 0.5793 - 1 \\
&= 1.1586 - 1 \\
&= 0.1586
\end{aligned}$$