Capitolo 5

Principali distribuzioni di probabilità continue

In questo capitolo presentiamo alcune distribuzioni di probabilità assolutamente continue.

5.1 La distribuzione uniforme

La distribuzione uniforme è la distribuzione di eventi equiprobabili. Nel caso continuo, una variabile aleatoria X si dice che segue una *distribuzione uniforme* nell'intervallo $[a, b]$, con $-\infty < a < b < \infty$, se ha la seguente funzione di densità:

$$f(x) := \begin{cases} \dfrac{1}{b-a} & \text{se } x \in [a, b] \\ 0 & \text{altrimenti.} \end{cases}$$

Scriveremo allora $X \sim U(a, b)$. La speranza matematica e la varianza di una variabile aleatoria $X \sim U(a, b)$ sono date rispettivamente da:

$$\mathbb{E}(X) = \frac{a+b}{2}$$
$$\mathrm{Var}(X) = \frac{(b-a)^2}{12} \ .$$

Figura 5.1: Funzione di densità di distribuzione uniforme continua su (a, b).

Il paradosso di Bertrand. Si sceglie a caso una corda di un cerchio di raggio R e centro C. Qual è la probabilità p che sia maggiore del lato del triangolo equilatero inscritto?

Il problema ha più di una soluzione in quanto non risulta chiaro il significato della frase "scegliere una corda a caso". Indichiamo con l la lunghezza del lato del triangolo equilatero inscritto.

1. Considerata una corda, sia A il punto intermedio. La corda è più lunga del lato del triangolo equilatero se il punto A ricade all'interno del cerchio di raggio $R/2$ e centro C. Il rapporto fra tali aree è 1/4 e pertanto risulta $p = 1/4$.

2. In base alla simmetria, un estremo della corda può essere scelto sulla circonferenza del cerchio in corrispondenza di uno dei vertici del triangolo. I vertici del triangolo dividono la circonferenza in tre archi di uguale lunghezza e pertanto si ha $p = 1/3$.

3. Scelto un punto B a caso, uniformemente sul raggio del cerchio, consideriamo la corda perpendicolare a tale raggio passante per B. Allora la corda casuale ha lunghezza maggiore di di l se il punto B appartiene alla metà del raggio più vicina al centro. In base alla simmetria, non importa quale raggio viene scelto e pertanto si ha $p = 1/2$.

Spiegazione . Ciascuno dei tre metodi utilizza una diversa distribuzione uniforme: 1) sul cerchio; 2) sulla circonferenza; 3) sul raggio del cerchio. ♣

5.2 La distribuzione esponenziale negativa

Abbiamo visto in precedenza che, sotto opportune ipotesi, il numero di volte X in cui un certo evento A si verifica in un intervallo di tempo di lunghezza T può descriversi mediante una distribuzione di Poisson:

$$P(X = x) := \mathrm{e}^{-\lambda T} \frac{(\lambda T)^x}{x!} \qquad\qquad x = 0, 1, 2, \ldots$$

Se si indica con X la v.a. che rappresenta il tempo necessario affinchè l'evento A si manifesti per la prima volta, allora X è una v.a. continua. La distribuzione di X si chiama *distribuzione esponenziale negativa,* e scriveremo $X \sim \mathcal{E}(\lambda)$; la sua funzione di densità è data da:

$$f(x) := \begin{cases} \lambda\, \mathrm{e}^{-\lambda x} & \text{per } x > 0 \\ \\ 0 & \text{altrove.} \end{cases} \tag{5.1}$$

Si noti che la (5.1) può essere scritta sinteticamente come:

$$f(x) \quad = \quad \lambda\, \mathrm{e}^{-\lambda x}\, I_{(0,+\infty)}(x)\ .$$

La verifica della (4.3) porge:

$$\int_{-\infty}^{+\infty} \lambda\, \mathrm{e}^{-\lambda x}\, I_{(0,+\infty)}(x)\, \mathrm{d}x = \int_0^{+\infty} \lambda\, \mathrm{e}^{-\lambda x}\, \mathrm{d}x = \left[-\mathrm{e}^{-\lambda x}\right]_0^{+\infty} = 1\ .$$

La funzione di ripartizione è data da:

$$F(x) := \int_{-\infty}^{x} \lambda \mathrm{e}^{-\lambda t} I_{(0,+\infty)}(t)\,\mathrm{d}t = \int_{0}^{x} \mathrm{e}^{-\lambda t}\, d(\lambda t) = \left[-\mathrm{e}^{-\lambda t}\right]_0^x = 1 - \mathrm{e}^{-\lambda x}\ .$$

La funzione:

$$S(x) := P(X > x) = 1 - F(x) = \mathrm{e}^{-\lambda x}$$

viene chiamata *funzione di sopravvivenza*. Essa esprime la probabilità che la v.a. X assuma un valore superiore a x.

La speranza matematica, integrando per parti, è data da:

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{+\infty} x\lambda\, \mathrm{e}^{-\lambda x} I_{(0,+\infty)}(x)\,\mathrm{d}x = \int_{0}^{+\infty} \lambda x\, \mathrm{e}^{-\lambda x}\,\mathrm{d}x \\ &= \left[x(-\mathrm{e}^{-\lambda x})\right]_0^{+\infty} + \int_0^{+\infty} \mathrm{e}^{-\lambda x}\, dx = \frac{1}{\lambda}\ .\end{aligned}$$

In maniera analoga si ottiene:

$$\mathrm{Var}(X) = \frac{1}{\lambda^2}$$

e quindi la distribuzione esponenziale negativa è caratterizzata dal fatto che la speranza matematica è uguale allo scarto quadratico medio.

Al pari della distribuuzione geometrica, la distribuzione esponenziale negativa possiede la proprietà di *assenza di memoria*, infatti:

$$\begin{aligned}P(X \le t + \Delta t | X > t) &= \frac{P(X \le t + \Delta t, X > t)}{P(X > t)} = \frac{P(t < X \le t + \Delta t)}{P(X > t)} \\ &= \frac{\mathrm{e}^{-\lambda t} - \mathrm{e}^{-\lambda(t+\Delta t)}}{\mathrm{e}^{-\lambda t}} = 1 - \mathrm{e}^{-\lambda \Delta t} = P(X \le \Delta t)\end{aligned}$$

essendo $\Delta t \ge 0$.

Si noti che la distribuzione esponenziale negativa può essere anche interpretata, invece che come tempo d'attesa di un evento, anche come legge di densità dell'intervallo di tempo che separa gli istanti in cui si verificano due eventi aleatori in accordo ad una legge di Poisson.

Esercizio 5.1 Un sistema S è composto da due componenti in serie A e B, i cui tempi di durata fino al guasto sono quantità aleatorie indipendenti, denotate rispettivamente con X e Y. Il vettore aleatorio (X, Y) ha una densità di probabilità $f(x, y) = k\mathrm{e}^{-x-2y}$, per $x > 0$ e $y > 0$. Calcolare

1. Il valore della costante k;

2. La funzione di sopravvivenza $S(z) = P(Z > z)$ del tempo di guasto Z di S;

3. la previsione $\mathbb{E}(Z)$ di Z.

Soluzione. In base alle ipotesi, la funzione di densità di (X, Y) è data da:

$$f(x,y) = \begin{cases} k\mathrm{e}^{-x-2y} & \text{per } x > 0, y > 0 \\ 0 & \text{altrimenti.} \end{cases}$$

In base alle ipotesi deve risultare:

$$\int_0^\infty \int_0^\infty f(x,y)\, dxdy = 1$$

cioè:

$$\int_0^\infty \int_0^\infty \mathrm{e}^{-x-2y}\, dxdy = \frac{1}{k}$$

da cui

$$\int_0^\infty \mathrm{e}^{-x}\, dx \int_0^\infty \mathrm{e}^{-2y}\, dy = 1 \cdot \frac{1}{2} = \frac{1}{k}$$

e quindi $k = 2$. La v.a. Z è data da $Z = \min(X, Y)$ e quindi:

$$P\{Z > z\} = P\{X > z\}P\{Y > z\} = \int_z^\infty \mathrm{e}^{-x}\, dx \int_z^\infty \mathrm{e}^{-2y}\, dy = \mathrm{e}^{-z} \cdot \mathrm{e}^{-2z} = \mathrm{e}^{-3z} \ ,$$

e quindi Z segue una distribuzione esponenziale di parametro $\lambda = 3$, cioè $Z \sim \mathcal{E}(3)$, e quindi $\mathbb{E}[Z] = \frac{1}{3}$. ♣

Proposizione 5.2 Siano X e Y due v.a. indipendenti aventi distribuzione esponenziale di parametro λ. Allora $Z = X/(X + Y)$ segue una distribuzione $U(0, 1)$.

5.3 La distribuzione normale

La distribuzione normale o gaussiana è la distribuzione che più di ogni altra trova applicazione in statistica. La ragione principale di ciò risiede nel fatto che essa costituisce un modello che approssima numerose altre distribuzioni e possiede notevoli proprietà matematiche.

Distribuzione normale standard. Diremo che una variabile aleatoria Z segue una *distribuzione normale standard* o gaussiana standard, e scriveremo $Z \sim N(1, 0)$, se ha la seguente funzione di densità:

$$f(z) := \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-z^2/2} \qquad z \in \mathbb{R} \ . \tag{5.2}$$

Per dimostrare che effettivamente la (5.2) è effettivamente una funzione di densità, dobbiamo dimostrare che vale la proprietà (4.3). A tal scopo, dimostriamo la seguente relazione:

$$\int_{-\infty}^{+\infty} \mathrm{e}^{-z^2/2}\, dz = \sqrt{2\pi} \ . \tag{5.3}$$

Consideriamo la quantità:

$$\left(\int_{-\infty}^{+\infty} \mathrm{e}^{-z^2/2}\,dz\right)^2 = \int_{-\infty}^{+\infty} \mathrm{e}^{-x^2/2}\,dx \cdot \int_{-\infty}^{+\infty} \mathrm{e}^{-y^2/2}\,dy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{e}^{(x^2+y^2)/2}\,dx\,dy\ .$$

Considerata la trasformazione in coordinate polari $(x, y) \to (\rho, \theta)$ definita da $x = \rho\cos\theta$ e $y = \rho\sin\theta$ e considerato che il modulo dello jacobiano[1] è uguale a ρ, si ottiene:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{e}^{(x^2+y^2)/2}\,dx\,dy = \int_{0}^{+\infty}\int_{0}^{2\pi} \mathrm{e}^{-\rho^2/2}\rho\,d\rho\,d\theta = 2\pi\int_{0}^{+\infty} \mathrm{e}^{-\rho^2/2}\rho\,d\rho = 2\pi\ .$$

Si ha quindi la (5.3), da cui segue:

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz = 1\ .$$

La speranza $\mathbb{E}[Z]$ è data da:

$$\begin{aligned}
\mathbb{E}[Z] &= \int_{-\infty}^{\infty} z f(z)\,dz = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z e^{-z^2/2}\,dz \\
&= \frac{1}{\sqrt{2\pi}}\left(-\int_{-\infty}^{0} e^{-z^2/2} d(\frac{z^2}{2}) - \int_{0}^{+\infty} e^{-z^2/2} d(\frac{z^2}{2})\right) \\
&= \frac{1}{\sqrt{2\pi}}\left(\left[-e^{-z^2/2}\right]_{-\infty}^{0} - \left[-e^{-z^2/2}\right]_{0}^{+\infty}\right) = \frac{1}{\sqrt{2\pi}}(-1+1) = 0\ .
\end{aligned}$$

Per quanto riguarda la varianza, poichè $\mathbb{E}[Z] = 0$, allora segue Var$(Z) = \mathbb{E}[Z^2]$ e, in base al teorema di integrazione per parti si ha:

$$\begin{aligned}
\mathrm{Var}(Z) = \mathbb{E}[Z^2] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2 e^{-z^2/2}\,dz = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z\cdot(z e^{-z^2/2})\,dz \\
&= \left[-z\cdot e^{-z^2/2}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-z^2/2}\,dz = 1\ .
\end{aligned}$$

La funzione di ripartizione della distribuzione normale standard viene usualmente denotata col simbolo $\Phi(z)$, cioè

$$\Phi(z) := P(Z \le z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} \mathrm{e}^{-t^2/2}\,dt\ .$$

Si noti che, per la simmetria della distribuzione normale standard, si ha:

$$\Phi(-z) = 1 - \Phi(z) \qquad \text{per ogni } z \in \mathbb{R}\ .$$

In particolare, dalla relazione precedente si ottiene la seguente relazione fra i quantili α e $(1-\alpha)$, con $0 < \alpha < 1$ di una distribuzione normale standard:

$$z_\alpha = -z_{1-\alpha} \qquad 0 < \alpha < 1\ , \tag{5.4}$$

che risulta molto utile nel calcolo pratico dei quantili di una distribuzione normale standard.

[1] Per il calcolo dello Jacobiano, si rimanda ai testi di analisi matematica.

Famiglia delle distribuzioni normali. Sia $Z \sim N(0,1)$ e siano $\mu \in \mathbb{R}$ e $\sigma > 0$. Consideriamo la v.a.

$$X := \sigma Z + \mu \; . \tag{5.5}$$

In base alla (4.24) la v.a. X ha funzione di densità:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad x \in \mathbb{R} \; . \tag{5.6}$$

In questo caso diremo che la v.a. X segue una *distribuzione normale* di parametri μ e σ e scriveremo $X \sim N(\mu, \sigma^2)$. Osserviamo che per $\mu = 0$ e $\sigma^2 = 1$ dalla (5.6) si ottiene la densità della normale standard (5.2)

In base alla (4.24), dalla (5.5), si ottiene la funzione di ripartizione di X:

$$F_X(x) := P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \; . \tag{5.7}$$

La (5.7) è equivalente alla seguente relazione fra i quantili di ordine α delle due distribuzioni normali $N(\mu, \sigma^2)$ e $N(0,1)$:

$$x_\alpha = \sigma z_\alpha + \mu \tag{5.8}$$

dove x_α e z_α denotano il quantile di ordine α rispettivamente della distribuzione $N(\mu, \sigma^2)$ e della distribuzione normale standard $N(0,1)$.

Con ragionamenti analoghi al caso standard, otteniamo la speranza matematica e la varianza di una variabile aleatoria $X \sim N(\mu, \sigma^2)$:

$$\mathbb{E}(X) = \mu \; ,$$
$$\text{Var}(X) = \sigma^2 \; .$$

Proposizione 5.3 Siano $X_1, \ldots, X_n$ v.a. indipendenti con $X_k \sim N(\mu_k, \sigma_k^2)$, per $k = 1, \ldots, n$. Allora il v.a. $X_1 + \cdots + X_n$ segue una distribuzione $N(\sum_{k=1}^n \mu_k, \sum_{k=1}^n \sigma_k^2)$.

Più in generale se $X_1, \ldots, X_n$ sono v.a. indipendenti aventi distribuzione rispettivamente $N(\mu_k, \sigma_k^2)$, con $k = 1, \ldots, n$ e $\alpha_1, \ldots, \alpha_n$ sono numeri reali, allora la variabile aleatoria $\sum_{k=1}^n \alpha_k X_k$ ha distribuzione $N(\sum_{k=1}^n \alpha_k \mu_k, \sum_{k=1}^n \alpha_k^2 \sigma_k^2)$.

Proposizione 5.4 Siano X, Y v.a. indipendenti con distribuzione $N(\mu_1, \sigma_1^2)$ e $N(\mu_2, \sigma_2^2)$. Allora $X + Y$ e $X - Y$ sono indipendenti.

Esercizio 5.5 Sia X una v.a. avente distribuzione normale con media $\mu = 10$ e varianza $\sigma^2 = 16$. Calcolare

a) $P(X > 12)$;

b) $P(7.5 \leq X \leq 11.2)$;

c) $P(|X - 10| < 2)$;

d) $P(X > 12 | X > 9)$.

Soluzione. Essendo $X \sim N(10, 10)$, per i punti a) e b) si ha rispettivamente:

$$\begin{aligned} P(X > 12) &= P\left(\frac{X-10}{4} > \frac{12-10}{4}\right) = P(Z > 0.5) = 1 - \Phi(0.5) \\ &= 1 - 0.6915 = 0.3085 \\ P(7.5 \le X \le 11.2) &= P\left(\frac{7.5-10}{4} \le \frac{X-10}{4} \le \frac{11.2-10}{4}\right) \\ &= P(-0.625 \le Z \le 0.3) = \Phi(0.3) - \Phi(-0.625) \\ &= \Phi(0.3) - 1 + \Phi(0.625) = 0.6179 - 1 + 0.7340 = 0.3519\,. \end{aligned}$$

dove Z denota la v.a. normale standard. Per quanto riguarda il terzo punto, osserviamo preliminarmente che la relazione $|X - 10| < 2$ è equivalente a

$$-2 < X - 10 < 2 \qquad \text{da cui segue:} \quad 8 < X < 12$$

e quindi:

$$\begin{aligned} P(|X-10| < 2) &= P(8 < X < 12)P\left(\frac{8-10}{4} \le \frac{X-10}{4} \le \frac{12-10}{4}\right) \\ &= P(-0.5 \le Z \le 0.5) = \Phi(0.5) - \Phi(-0.5) \\ &= \Phi(0.5) - 1 + \Phi(0.5) = 0.6915 - 1 + 0.6915 = 0.3829\,. \end{aligned}$$

Infine, tenendo conto che

$$P(X > 12|X > 9) = \frac{P(\{X > 12\} \cap \{X > 9\})}{P(X > 9)} = \frac{P(X > 12)}{P(X > 9)}$$

ed essendo, in base al punto a) $P(X > 12) = 0.3085$, dobbiamo ricavare preliminarmente solo $P(X > 9)$

$$\begin{aligned} P(X > 9) &= P\left(\frac{9-10}{4} > \frac{9-10}{4}\right) = P(Z > -0.25) = 1 - \Phi(0.25) \\ &= 1 - 0.5987 = 0.4013 \end{aligned}$$

da cui segue

$$P(X > 12|X > 9) = \frac{P(X > 12)}{P(X > 9)}) = \frac{0.3085}{0.4013} = 0.7688\,.$$

♣

Esercizio 5.6 Sia X una v.a. normale avente speranza matematica $\mu = 4$. Calcolare:

1. la varianza σ^2 di X sapendo che $P(X > 5) = 0.3707$;
2. $P(|X - 5| < 2)$;
3. $P(X > 7|X > 4)$.

Soluzione. Per quanto riguarda il primo punto, dall'equazione $P(X > 5) = 0.3707$ segue:

$$P(X > 5) = P\left(\frac{X-4}{\sigma} > \frac{5-4}{\sigma}\right) = P\left(Z > \frac{1}{\sigma}\right) = 0.3707$$

da cui:

$$\Phi\left(\frac{1}{\sigma}\right) = 1 - \Phi\left(\frac{1}{\sigma}\right) = 1 - 0.3707 = 0.6293\,.$$

Indicato con $z_{0.6293}$ il quantile 0.6293 (cioè il valore per cui la funzione di ripartizione assume valore 0.6293), dalle tavole si ottiene $z_{0.6293} = 0.33$ da cui:

$$z_{0.6293} = \frac{1}{\sigma} = 0.33$$

e quindi $\sigma = 3$. Pertanto $X \sim N(4, 9)$.

Per quanto concerne il secondo punto, preliminarmente conviene scrivere l'insieme $\{|X - 5| < 2\}$ come $\{-2 < X - 5 < 2\}$ cioè $\{3 < X < 7\}$. Si ha quindi:

$$\begin{aligned} P(3 < X < 7) &= P\left(\frac{3-4}{3} < \frac{X-4}{3} < \frac{7-4}{3}\right) = P\left(-\frac{1}{3} < Z < 1\right) = \Phi(1) - \Phi\left(-\frac{1}{3}\right) \\ &= \Phi(1) - [1 - \Phi\left(\frac{1}{3}\right)] = \Phi(1) + \Phi\left(\frac{1}{3}\right) - 1 \\ &= 0.8413 + 0.6293 - 1 = 0.4706\,. \end{aligned}$$

Infine si ha:

$$\begin{aligned} P(X > 7|X > 4) &= \frac{P(\{X > 7\}, \{X > 4\})}{P(X > 4)} = \frac{P(X > 7)}{0.5} = 2 \cdot P\left(\frac{X-4}{3} > \frac{7-4}{3}\right) \\ &= 2[1 - \Phi(1)] = 2[1 - 0.8413] = 0.3174 \end{aligned}$$

♣

Esercizio 5.7 Un'industria produce su commissione delle sbarre di acciaio cilindriche, il cui diametro deve essere di 4 cm, ma che tuttavia sono accettabili se hanno un diametro compreso fra 3.995 cm e 4.005 cm. Il cliente, nel controllare le sbarre fornitegli, constata che il 5% sono di diametro inferiore a quello tollerato ed il 12% sono di diametro superiore. Supponendo che le misure x dei diametri seguano una distribuzione normale, calcolarne i parametri μ e σ^2.

Determinare inoltre il valore di σ affinchè la probabilità che le sbarre abbiano un diametro superiore a quello tollerato sia inferiore al 2%.

Soluzione. Valutando la probabilità con la frequenza, in base ai dati del problema possiamo impostare il sistema:

$$\begin{aligned} P\{X < 3.995\} &= 0.05 \qquad (5.9) \\ P\{X > 4.005\} &= 0.12 \end{aligned}$$

In particolare l'ultima relazione, in termini di funzione di ripartizione si scrive equivalentemente:

$$P\{X \leq 4.005\} = 1 - 0.12 = 0.88\ . \tag{5.10}$$

Al fine della risoluzione del problema, possiamo riscrivere le relazioni (5.9) e (5.10) in termini di quantili:

$$\begin{aligned} x_{0.05} &= 3.995 \\ x_{0.88} &= 4.005 \end{aligned}$$

e quindi, tenendo conto della (5.8) si ha:

$$\begin{aligned} x_{0.05} &= \sigma z_{0.05} + \mu = 3.995 \\ x_{0.88} &= \sigma z_{0.88} + \mu = 4.005\ . \end{aligned}$$

Le quantità $z_{0.05} = -1.65$ e $z_{0.88} = 1.18$ sono rispettivamente i quantili 0.05 e 0.88 della distribuzione normale standard, ottenuti dalle tavole della stessa distribuzione. Otteniamo pertanto il sistema:

$$\begin{aligned} -1.65\sigma + \mu =& 3.995 \\ 1.18\sigma + \mu =& 4.005 \end{aligned}$$

da cui si ottengono i valori: $\mu = 4.001$ e $\sigma = 0.0035$.

Per quanto riguarda l'ultimo punto, si richiede di calcolare il valore dello scarto σ tale che:

$$P(X > 4.005) < 0.02$$

da cui riscrivendo in termini di funzione di ripartizione di X:

$$P(X \leq 4.005) = 1 - P(X > 4.005) \geq 0.98$$

e quindi, in altre parole, il valore $x = 4.005$ coincide con il quantile 0.98, cioè $x_{0.98} = 4.005$. Dalla nota relazione $x_{0.98} = \sigma \cdot z_{0.98} + \mu$ che lega i quantili di una distribuzione normale di media μ e varianza σ^2 ai corrispondenti quantili di una distribuzione standard, utilizzando il valore della media $\mu = 4.001$ ottenuto in precedenza e e ricavando dalle tavole il quantile 0.98 della variabile Z, $z_{0.98} = 2.06$, ricaviamo infine:

$$4.005 = \sigma \cdot 2.06 + 4.001 \qquad \frac{0.004}{2.06} = \sigma\ .$$

da cui $\sigma = 0.0019$. Tale valore corrisponde ad una percentuale di sbarre aventi un diametro superiore a quello tollerato pari al 2%; per valori di σ inferiori a 0.0019 si ottiene ovviamente percentuali inferiori di prodotto non conforme alle specifiche. ♣

Approssimazione della distribuzione binomiale mediante la distribuzione normale. Sia $X \sim B(n,p)$. Per valori grandi di n (tipicamente superiori a 50), la distribuzione di X può essere approssimata mediante una distribuzione normale avente speranza e varianza rispettivamente uguali a np e npq cioè uguali ad i rispettivi valori della distribuzione binomiale. In altre parole, per $n \geq 50$ si può assumere:

$$B(n,p) \approx N(np, npq) \ .$$

Il risultato teorico che giustifica questa approssimazione è il *Teorema di De Moivre-Laplace* che costituisce un'applicazione del noto *Teorema del Limite Centrale*.

Per l'applicazione pratica di tale risultato, deve essere prevista una *correzione di continuità*, per cui:

$$\begin{aligned} P(X \leq x) &= P(X < x + \frac{1}{2}) = P\left(\frac{X-np}{\sqrt{npq}} < \frac{x+\frac{1}{2}-np}{\sqrt{npq}}\right) \\ &= P\left(Z < \frac{x+\frac{1}{2}-np}{\sqrt{npq}}\right) \simeq \Phi\left(\frac{x+\frac{1}{2}-np}{\sqrt{npq}}\right) \end{aligned}$$

Esercizio 5.8 Su 1000 nascite, qual è la probabilità che la percentuale di nati maschi si trovi fra il 48% ed il 52%?

Si indichi con X la v.a. "nato maschio" e si supponga che ogni nascita sia indipendente, nel senso che il sesso dei nati precedentemente non influenzi quello dei nati successivamente. Assumiamo inoltre, per seplicità, il valore $p = 0.5$. In base alle assunzioni si ha $X \sim B(1000, 0.5)$. Possiamo approssimare la distribuzione binomiale mediante una distribuzione normale di parametri $\mu = np = 500$ e scarto $\sigma = \sqrt{npq} = 15.8$. Pertanto

$$\begin{aligned} P(480 \leq X \leq 520) &= P(480 + \frac{1}{2} < X < 520 + \frac{1}{2}) \\ &= P\left(\frac{480+\frac{1}{2}-500}{15,8} < \frac{X-500}{15,8} < \frac{520+\frac{1}{2}-500}{15,8}\right) \\ &= P(-1,297 < Z < 1,297) = \Phi(1,297) - \Phi(-1,297) \\ &= 2\Phi(1,297) - 1 = 0.8064 \ . \end{aligned}$$

Pertanto la probabilità che i nati maschi siano compresi fra il 48% e il 52% è all'incirca del 80.64%. ♣

5.4 Altri esercizi

Esercizio 5.9 Siano X_1, X_2, X_3, X_4 quattro variabili aleatorie indipendenti ed identicamente distribuite con legge uniforme nell'intervallo $[0,2]$. Considerata la variabile aleatoria

$$Z = \max\{X_1, X_2, X_3, X_4\}$$

calcolare la densità di probabilità ed il valore atteso della v.a. $Y = 2 - Z$.

Esercizio 5.10 Sia X una v.a. continua avente distribuzione uniforme sull'intervallo $[a, b]$, con $0 < a < b$. E' noto che la speranza e la varianza di X sono uguali rispettivamente a $\mathbb{E}(X) = 6$ e Var$(X) = 3$. Calcolare $P(X \geq 7)$ e $P(2.3 \leq X \leq 9.4)$.

Esercizio 5.11 Un sistema S è composto da 4 componenti A, B, C e D. Il componente A è in serie con B, mentre C è in serie con D. Inoltre A e B sono in parallelo con C e D. I tempi (aleatori) fino al guasto dei quattro componenti sono rispettivamente X, Y, Z e W. Il vettore aleatorio (X, Y, Z, W) ha come densità la funzione:

$$f(x, y, z, w) = \begin{cases} \mathrm{e}^{-(x+y+z+w)} & \text{per } x \geq 0, y \geq 0, z \geq 0, w \geq 0 \\ 0 & \text{altrove.} \end{cases}$$

Calcolare:

1. la densità congiunta $g(x, y)$ del vettore (X, Y) per ogni $x \geq 0$, $y \geq 0$;
2. il valore atteso μ del tempo di guasto del sottosistema costituito dai componenti A e B in serie;
3. la probabilità τ che il sistema S non si guasti fino ad un determinato tempo t.

Esercizio 5.12 Si considerino una variabile aleatoria avente distribuzione uniforme sull'insieme $\{-2, 2\}$ ed una v.a. Y, indipendente da X, avente distribuzione normale standard. Posto $Z = XY + 1$, determinare:

1. il coefficiente di correlazione ρ di X e Z;
2. la funzione di ripartizione $F(z)$;
3. il valore k tale che $P(Z > 2k + 1) = 0,1587$.

Esercizio 5.13 Un impiegato abita all'estrema periferia della città e la strada più breve per raggiungere il suo ufficio passa per il centro, in questo caso il tempo richiesto per percorrere tragitto può essere descritto mediante una variabile aleatoria T_1 di media $\mu_1 = 28$ minuti e scarto $\sigma_1 = 7$ minuti e 20 secondi. Sono possibili altre due strade alternative il cui tragitto può essere descritto mediante rispettivamente una v.a. T_2 con media $\mu_2 = 32$ minuti e scarto $\sigma_2 = 4$ minuti e 30 secondi ed una v.a. T_3 con media $\mu_3 = 33$ minuti e scarto $\sigma_3 = 1$ minuto e 10 secondi. Assumendo che le tre v.a. abbiano tutte distribuzione normale, calcolare quale sia tragitto più opportuno da percorrere nel caso in cui: *i.*) l'impiegato debba essere in ufficio entro 30 minuti; *ii.*) l'impiegato debba essere in ufficio entro 35 minuti?

Esercizio 5.14 Sia (X, Y) una v.a. doppia a valori in $\{(x, y) \in \mathbb{R}^2 \; : \; 0 \leq x \leq 1 \, , 0 \leq y \leq x/3\}$ ed avente densità:

$$f(x, y) = \begin{cases} c(x - 3y) & \text{per } 0 \leq x \leq 1 \, , 0 \leq y \leq x/3 \\ 0 & \text{altrove.} \end{cases}$$

Calcolare il valore della costante c e le funzioni di densità $f_1(x)$ e $f(y|x)$.

Esercizio 5.15 Sia X una v.a. continua avente funzione di distribuzione F_X e funzione di densità f_X. Dimostrare che la variabile aleatoria $Y = X^2$ è ancora continua ed esprimere la sua funzione di distribuzione F_Y e la sua funzione di densità f_Y in termini di F_X e f_X. Inoltre calcolare la funzione di densità di $Y = X^2$ nei casi in cui:

1. X segue una distribuzione normale $N(\mu, \sigma^2)$;
2. X segue una distribuzione di Laplace avente densità $f(x) = \frac{1}{2}\mathrm{e}^{-|x|}$;
3. X segue una distribuzione di Cauchy avente densità $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.

Esercizio 5.16 Siano X_1 e X_2 due v.a. indipendenti aventi distribuzione normale con medie rispettivamente $\mu_1 = 10$, $\mu_2 = 4$ ed aventi la stessa varianza $\sigma^2 = 9$.

a) $P(X_1 + X_2 > 16)$;

b) $P(2X_1 > 16)$;

c) $P(7 < X_1 - X_2 < 10)$;

d) $P(X_2 - X_1 > -5)$.

Esercizio 5.17 Si consideri un negozio in cui arrivano, a caso, mediamente 20 clienti l'ora.

1. Calcolare la probabilità che gli intervalli di tempo fra due arrivi consecutivi siano inferiori a tre minuti;
2. Calcolare la probabilità che gli intervalli di tempo fra due arrivi consecutivi siano superiori a quattro minuti;
3. Supponendo che il 10% dei clienti compri un certo oggetto, ricavare la distribuzione di probabilità del numero di cliento che acquistano un oggetto in un'ora.

Esercizio 5.18 Siano $f_1(x)$ e $f_2(x)$ due funzioni di densità esponenziali di parametro rispettivamente $\lambda_1 = 1$ e $\lambda_2 = 2$. Si consideri la funzione $f(x)$ definita come segue:

$$f(x) = \alpha f_1(x) + (1-\alpha) f_2(x) \qquad 0 < \alpha < 1 \; .$$

1. Dimostrare che la funzione $f(x)$ è anch'essa una funzione di densità.
2. Calcolare la speranza matematica di $f(x)$.
3. Denotata con X una v.a. avente legge $f(x)$, calcolare il valore della costante α tale che $P(X > 2) = 0.1$.

Esercizio 5.19 Sia X la vita in ore di un certo tipo di lampadine. Da osservazioni passate, è noto che X segue una distribuzione normale $N(200, \sigma^2)$. Se un acquirente vuole che almeno il 95% di una partita di tali lampadine abbia una vita superiore alle 150 ore, calcolare quale può essere la massima varianza σ^2 della popolazione di lampadine che il compratore sia disposto ad accettare.

Esercizio 5.20 Determinare i parametri della distribuzione di una variabile aleatoria $X \sim N(\mu, \sigma^2)$ sapendo che il primo quartile di X è uguale a 40 ed il terzo quartile è uguale a 60. Successivamente calcolare $P(X < 55|X > 40)$.

Esercizio 5.21 Siano X_1 e X_2 due v.a. aventi distribuzione binomiale di parametri rispettivamente $n_1 = 350$ e $p_1 = 0.3$, $n_2 = 50$, $p_2 = 0.3$, cioè $X_1 \sim B(350, 0.3)$, $X_2 \sim B(50, 0.3)$. Considerata la v.a. $Y = X_1 - X_2$, calcolare

1. la mediana e la varianza di di Y;
2. $P(Y > 90)$.

Esercizio 5.22 Siano $X \sim N(10, 16)$ e $Y \sim N(16, 9)$ v.a. indipendenti.

i) Rappresentare graficamente le due curve normali su uno stesso diagramma,

Successivamente calcolare:

ii) $P(8 < X \leq 11)$,

iii) $P(|X - 8| < 2)$,

iv) il quantile 0.3707 di X, cioè $x_{0.3707}$,

v) $P(X + Y > 30)$,

vi) $P(2X - Y < 3)$.

Esercizio 5.23 Siano $X \sim B(10, 0.1)$ e $Y \sim B(90, 0.1)$ v.a. indipendenti. Calcolare: Successivamente calcolare:

ii) $P(X \geq 2)$,

iii) $P(Y \leq 10)$,

iv) $P(5 < X + Y < 20)$,

Esercizio 5.24 Si supponga che il peso in grammi di una confezione di tonno in scatola (peso netto sgocciolato) segua una distribuzione normale. Sapendo che il 3% delle confezioni prodotte ha peso inferiore ai 75.18 grammi e che il 2% delle confezioni di tonno ha un peso superiore ai 81.08 grammi, calcolare la media e la varianza della distribuzione normale in esame.

Esercizio 5.25 Siano $X \sim N(20, 25)$ e $Y \sim N(10, 16)$ v.a. indipendenti.

i) Rappresentare graficamente le due curve normali su uno stesso diagramma,

Successivamente calcolare:

ii) $P(18 < X \leq 23)$,

iii) $P(|X - 18| < 2)$,

iv) il quantile 0.3707 di X, cioè $x_{0.3707}$,

v) $P(X + Y > 30)$,

vi) $P(X - 3Y < -5)$.

Esercizio 5.26 Siano $X \sim \mathcal{B}(15, p)$ e $Y \sim \mathcal{B}(135, p)$ v.a. indipendenti.

i) Calcolare p sapendo che $\mathbb{E}(X) = 1.5$ e Var$(X) = 1.35$

Successivamente calcolare:

ii) $P(X \leq 1)$,

iii) $P(X \geq 14)$

iv) $P(10 \leq X + Y \leq 20)$,

Esercizio 5.27 Un'azienda alimentare propone un nuovo tipo di merenda al cioccolato e crema di zabaione, rivolta particolarmente ai ragazzi aventi un'età compresa fra 10 e 12 anni. L'area marketing ha successivamente rilevato che, a livello nazionale, il 30% degli interessati ha visto almeno uno degli spot televisivi predisposti per la campagna pubblicitaria che ha lanciato il prodotto. Considerato un gruppo di 150 potenziali consumatori, ci si chiede:

1. qual è il numero atteso che ha visto almeno uno spot televisivo?

2. qual è la probabilità che ve siano stati almeno 50 che ha visto almeno uno spot?

Esercizio 5.28 La resistenza a trazione della carta usata per i sacchetti per alimenti è una caratteristica significativa della loro qualità. Si assuma che tale grandezza si distribuisca con legge normale di parametri $\mu = 2.8$ Kg/cm^2 e scarto quadratico medio $\sigma = 0.14$ Kg/cm^2. Se un'acquirente richiede sacchetti aventi una resistenza alla trazione almeno pari a 2.60 Kg/cm^2, calcolare la percentuale di sacchetti che soddisfano tale specifica. Successivamente calcolare il minimo valore di resistenza alla trazione che viene assicurato nel 95% dei sacchetti prodotti dall'azienda.