Esercitazioni di Statistica Matematica A
Lezione 2

Variabili con distribuzione gaussiana

1.1) Una bilancia difettosa ha un errore sistematico di 0.1g ed un errore casuale che si suppone avere la distribuzione $\mathcal{N}(0, 4/25)$ (si osservi che si possono avere anche risultati negativi!). Si sottopone alla misura un campione di 0.9g. Stimare:
a) la probabilità che la misura dia un risultato compreso tra 0.8g e 1g;
b) la probabilità che la misura dia un risultato superiore a 1g;
c) la probabilità che il valore assoluto della misura sia minore di 0.9g.

1.2) L'altezza di una data popolazione maschile di età compresa tra i 18 ed i 25 anni segue una distribuzione normale.
a) Se l'altezza media fosse 165cm ed esattamente il 5.48% della popolazione fosse più basso di 158cm, quale percentuale della popolazione avrebbe un'altezza superiore a 174cm?
b) Se il primo quartile dell'altezza fosse 160cm ed il 90^o percentile fosse 184cm, quanto varrebbero la media e la deviazione standard?

1.3) Il tempo necessario a Gianni per coprire il percorso casa-ufficio è una variabile aleatoria di legge normale. Se il tempo medio è di 30 minuti e la probabilità di coprire il percorso in più di 40 minuti è 0.1, quanto vale la probabilità di coprire il percorso in più di 50 minuti?

1.4) La durata in ore di una lampadina segue una legge normale di media 2000 e varianza σ^2. Se un compratore richiede che almeno il 90% di esse abbia una durata superiore alle 1500 ore, qual'è il valore massimo che σ può assumere per soddisfare l'esigenza dell'acquirente?

1.5) Una fabbrica che fornisce le FS costruisce rotaie la cui lunghezza segue una legge normale $\mathcal{N}(9.99, 1/10000)$ (dove la media è in metri e la varianza in metri^2). Determinare:
a) la probabilità che una rotaia sia più lunga di 10m;
b) la probabilità che 100 rotaie siano più lunghe di 1Km;
c) la probabilità che 101 rotaie siano più lunghe di 1Km;
d) la probabilità che esattamente 101 rotaie siano necessarie per coprire 1km;
e) la lunghezza x affinché la percentuale di rotaie con lunghezza non superiore a x sia il 10%;
f) Come si deve modificare la lunghezza media affinché la percentuale di rotaie con lunghezza superiore a 10m sia il 40%.

1.6) Un operatore, in attesa dell'apertura del mercato azionario, decide di comperare un pacchetto di azioni se la differenza tra il prezzo all'apertura e quello alla chiusura della sera precedente sarà compreso tra a e b ($a < b \in \mathbb{R}$). Supponendo che la variazione di prezzo segua una legge normale $\mathcal{N}(0, \sigma^2)$, qual'è il valore di σ in corrispondenza al quale si ha la massima probabilità di comperare?

Variabili con distribuzione Esponenziale e Gamma

2.1) Supponiamo che la durata della vita di ogni membro di un gruppo segua una legge esponenziale di parametro $\lambda = 1/50$ (in anni). Per un membro di tale popolazione si calcoli:

a) la probabilità di arrivare alla pensione (65 anni);
b) la probabilità di arrivare a 70 anni se ha appena compiuto il 40^o compleanno;
c) per quale valore di x si ha che $\mathbb{P}(T > x) = 1/2$ (dove T è la durata di vita dell'individuo considerato)?

2.2) Una variabile X segue la legge esponenziale di parametro λ. Determinare la funzione di ripartizione di X^2 e la sua eventuale densità.

2.3) Uno speciale dispositivo messo a punto in un laboratorio, permette di mandare una particella ad una distanza X che segue una variabile esponenziale di parametro λ, il quale può essere variato. Come si deve scegliere λ affinchè sia massima la probabilità di colpire un bersaglio di lunghezza l posto ad una distanza d?

2.4) Un componente elettronico è soggetto a guasti che seguono un processo di Poisson. Se mediamente si ha un guasto ogni 10 settimane, si determini il tempo medio di attesa per il 10^o guasto. Dare una formula analoga per il guasto n-esimo.

2.5) Tre persone (A, B e C) arrivano contemporaneamente all'ufficio postale. Le durate del servizio allo sportello per ciascun individuo si suppongano variabili indipendenti di legge esponenziale con parametro λ. Supponiamo che A e B occupino subito i due sportelli disponibili e che C si presenti al primo dei due sportelli che si libererà, calcolare:

a) la probabilità che C non sia l'ultimo a lasciare la posta;
b) la funzione di ripartizione ed il valore atteso del tempo totale speso da C nell'ufficio postale.

Variabili continue con altre distribuzioni

3.1) Sia X una variabile con distribuzione uniforme sull'intervallo $(0, 1)$. Trovare la funzione di ripartizione e l'eventuale densità di:

a) $Y = X^2$;
b) $Y = 1/X$;
c) $Y = \exp(X)$.

3.2 Siano X e Y indipendenti ed identicamente distribuite con distribuzione uniforme sull'intervallo $(0, 1)$. Se $X_1 := \min(X, Y)$, $X_2 := \max(X, Y)$, calcolare la funzione di ripartizione e le eventuali densità di X_1 e X_2. Si calcolino inoltre medie e varianze.

3.3) Sia X con distribuzione avente densità

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2} \qquad \text{(Cauchy).}$$

Calcolare $\mathbb{P}(X \geq 1)$ e $\mathbb{P}(|X| \geq 1)$.

Soluzioni

1.1)

a) Essendo $X \sim \mathcal{N}(0, 4/25)$ allora $5X/2 \sim \mathcal{N}(0,1)$; inoltre $M = X+1 \in [4/5, 1]$ se e solo se $X \in [-1/5, 0]$, pertanto $\mathbb{P}(X \in [-1/5, 0]) = \mathbb{P}(5X/2 \in [-1/2, 0]) = \phi(1/2) - 1/2 \approx 0.1915$ (si ricordi che $\phi(x) + \phi(-x) = 1$).

b) $X + 1 \leq 1$ se e solo se $X \leq 0$ quindi $\mathbb{P}(X \leq 0) = 1/2$.

c) $|X+1| \leq 9/10$ se e solo se $X \in [-19/10, -1/10]$, pertanto $\mathbb{P}(X \in [-19/10, -1/10]) = \mathbb{P}(5X/2 \in [-19/4, -1/4]) = \phi(19/4) - \phi(1/4) \approx 1 - 0.5987 = 0.4013$.

1.2) Ricordiamo che $X \sim \mathcal{N}(\mu, \sigma^2)$ implica $(X-\mu)/\sigma \sim \mathcal{N}(0,1)$ e che, se U_X è la funzione dei quantili di X allora $U_X = \mu + \sigma q$ (dove q è la funzione dei quantili della distribuzione $\mathcal{N}(0,1)$. Si ricordi inoltre che $q(\alpha) = -q(1-\alpha)$.

a) Sappiamo che $\mu = 165$ e che $0.0548 = \mathbb{P}(X \leq 158) = \phi((158-\mu)/\sigma)) = \phi(-7/\sigma)$, pertanto $\sigma = -7/q(0.0548) = 7/q(0.9452) \approx 7/1.6 = 4.375$. Da cui $\mathbb{P}(X > 174) = 1 - \phi((174-\mu)/\sigma) = 1 - \phi(2.0571) = 1 - 0.9802 = 0.0198$.

b) Se $160 = U_X(1/4) = \mu + \sigma q(1/4)$, $184 = U_X(9/10) = \mu + \sigma q(9/10)$ allora $\sigma = (184 - 160)/(q(9/10) - q(1/4)) \approx 24/(1.2816 - 0.6745) = 12.2693$, mentre $\mu = (160q(9/10) - 184q(1/4))/(q(9/10) - q(1/4)) \approx 329.1544/1.956 = 168.2758$.

1.3) $1/10 = \mathbb{P}(X > 40) = 1 - \phi((40-30)/\sigma)$ pertanto $\sigma = 10/q(9/10) \approx 10/1.2816 = 7.8027$. Da cui $P(X > 50) \approx 1 - \phi(20/7.8027) = 1 - \phi(2.5632) = 1 - 0.9948 = 0.0052$.

1.4) Se X è la durata della lampadina, l'aquirente richiede $9/10 \leq \mathbb{P}(X \geq 1500) = \mathbb{P}((X-\mu)/\sigma \geq (1500-\mu)\sigma) = 1 - \phi(-500/\sigma)$. Pertanto si ottiene $\sigma \leq -500/q(1/10) = 500/q(9/10) \approx 500/1.2816 \approx 390.137$.

1.5) Sia X la misura di una rotaia.

a) $\mathbb{P}(X > 10) = 1 - \phi((10 - 9.99)/0.01) = 1 - \phi(1) \approx 1 - 0.8143 = 0.1857$.

b) Se $\{X_i\}_{i=1}^{100}$ sono le misure (indipendenti ed identicamente distribuite) delle 100 rotaie, allora $\sum_{i=1}^{100} X_i \sim \mathcal{N}(100 \cdot 9.99, 100 \cdot (1/10000))$, pertanto $\mathbb{P}(\sum_{i=1}^{100} X_i > 1000) = 1 - \phi(1/0.1) = 1 - \phi(10) \approx 0$.

c) Similmente $\mathbb{P}(\sum_{i=1}^{101} X_i > 1000) = 1 - \phi((1000 - 9.99 \cdot 101)/\sqrt(0.0101)) = 1 - \phi(-89.9/\sqrt(1.01)) \approx 1$.

d) Sebbene un calcolo esatto di questo tipo coinvolgerebbe una convoluzione, in questo specifico caso possiamo stimare la probabilità in questo modo: sia A:"le prime 100 rotaie non bastano a coprire 1Km" e B:"le 101 rotaie bastano a coprire 1Km". Quello che dobbiamo calcolare è $\mathbb{P}(A \cap B)$ cioè $P(A) - P(A \cap B^c) \geq P(A) - P(B^c) \approx 1$. (Si osservi che, essendo variabili normali, le lunghezze non sono necessariamente positive e quindi $B^c \not\subset A$, anche se nel nostro caso la probabilità dell'evento $B^c \cap A^c$ è prossima a 0).

e) $x = U_X(0.1) = \mu + \sigma q(0.1) = 9.99 - 0.01q(0.9) \approx 9.99 - 0.01 \cdot 1.2816 = 9.9772$.

f) Se $4/10 = \mathbb{P}(Y\sigma + \mu > 10)$ (dove $Y \sim \mathcal{N}(0,1)$ e $\sigma = 0.01$) allora, equivalentemente, $(10-\mu)/\sigma = q(6/10)$ cioè $\mu = 10 - 0.01q(6/10) \approx 10 - 0.002533 = 9.997467$.

1.6) Sia X la differenza di quotazione, $\mathbb{P}(X \in [a,b]) = f(\sigma)$, dove, dopo un cambio di variabile

$$f(\sigma) = \int_{a/\sigma}^{b/\sigma} \frac{\exp(-s^2/2)}{\sqrt{2\pi}} \mathrm{d}s.$$

Calcolando la derivata f' e studiando la funzione in $(0,+\infty)$ si trova immediatamente che:

a) se $a \leq 0 \leq b$ ($a < b$) allora f è decrescente, pertanto minore è σ, maggiore è la probabilità che $X \in [a,b]$ (ed il limite per σ che tende a 0^+ di tale valore è 1);

b) se $0 < a < b$ allora si trova che f è decrescente in $(\sqrt{(b^2-a^2)/(2\ln(b/a))}, +\infty)$ e crescente in $(0, \sqrt{(b^2-a^2)/(2\ln(b/a))})$, pertanto il massimo è assunto in $\sigma^2 = (b^2 - a^2)/(2\ln(b/a))$.

c) se $a < b < 0$, utilizzando la parità ed il punto precedente si ha che il massimo è assunto in $\sigma^2 = (a^2 - b^2)/(2\ln(a/b))$.

2.1) La funzione di ripartizione della legge esponenziale di parametro λ è $F_X(t) = (1-\exp(-\lambda t))1\!\!1_{(0,+\infty)}(t)$; nel nostro caso $\lambda = 1/50$.

a) $\mathbb{P}(X > 65) = 1 - F_X(65) = \exp(-13/10) \approx 0.2725$.

b) Per la proprietà di assenza di memoria della legge esponenziale $\mathbb{P}(X > 70 | X > 40) = P(X > 30) = \exp(-3/5) \approx 0.5488$.

c) Detta $U_X(s) := \inf\{x : F_X(x) \geq s\}$ ($s \in (0,1)$) la funzione dei quantili della variabile X, dalla continuità di F_X, si ha che $F_X(x) = 1/2$ se e solo se $x = U_X(1/2)$. Facilmente si calcola $U_X(s) = (1/\lambda)\ln(1/(1-\lambda))$, pertanto $x = 50\ln(2) \approx 34.6574$.

2.2) La funzione di ripartizione è

$$F_{X^2}(t) := \mathbb{P}(X^2 \leq t) = \begin{cases} 0 & t < 0 \\ \mathbb{P}(X \in [-\sqrt{(t)}, \sqrt{(t)}]) = 1 - \exp(-\lambda\sqrt{t}) & \text{t} \geq 0. \end{cases}$$

La funzione di ripartizione è continua e derivabile in $\mathbb{R} \setminus \{0\}$ con derivata integrabile in $(0,+\infty)$ e $(-\infty,0)$. Dal teorema fondamentale del calcolo si ha che la derivata (completata con un valore arbitrario in 0) è una densità

$$\rho_{X^2}(t) = \begin{cases} 0 & t \leq 0 \\ \frac{\lambda\exp(-\lambda\sqrt{t})}{2\sqrt{t}} & t > 0. \end{cases}$$

2.3) Sappiamo che $\mathbb{P}(X \in (d, d+l)) = \exp(-d/\lambda) - \exp(-(d+l)/\lambda) =: g(\lambda)$. Dallo studio della derivata

$$g'(\lambda) = \left[\frac{d}{\lambda^2} - \frac{d+l}{\lambda^2}\exp(-l/\lambda)\right]\exp(-d/\lambda)$$

si trova che il massimo è assunto in $\lambda = l/\ln(1 + l/d)$.

2.4) Sappiamo che il tempo di attesa T_n dell'evento n-esimo di un processo di Poisson di parametro λ segue una legge $\Gamma(n,\lambda)$ con densità

$$\rho_{\Gamma(n,\lambda)}(t) = \frac{t^{n-1}}{\Gamma(n)}\lambda^n \exp(-\lambda t)1\!\!1_{(0,+\infty)}(t)$$

con media n/λ e varianza n/λ^2. Essendo $\lambda = 1/10$ (in settimane), si ha che $E(T_n) = 10n$, da cui nel caso particolare $n = 10$ si ottiene il tempo medio di attesa del 10^o guasto (100 settimane).

2.5)

a) Detti T_a, T_B e T_C le durate del servizio allo sportello, si ha che la probabilità cercata è $\mathbb{P}(T_A \geq T_B + T_C) + \mathbb{P}(T_A \geq T_B + T_C) \equiv 2\mathbb{P}(T_A \geq T_B + T_C)$. Ricordando che la densità della legge esponenziale è $\rho(x) = \lambda \exp(-\lambda x)1\!\!1_{(0,+\infty)}$ e che la legge congiunta (T_A, T_B, T_C) ha densità $\rho(x)\rho(y)\rho(z)$, si calcola

$$\mathbb{P}(T_A \geq T_B + T_C) = \int_{\{x \geq y+z\}} \rho(x)\rho(y)\rho(z)\mathrm{d}x\mathrm{d}y\mathrm{d}z = \ldots = 1/4.$$

Quindi la probabilità cercata 'e pari a $1/2$ (Si osservi che tale risultato risulta piuttosto ovvio tenendo conto della proprietà di assenza di memoria della legge esponenziale).

b) La densità della legge $T_C + \min(T_A, T_B)$ si ottiene (per l'indipendenza degli addendi) come convoluzione delle due densità. Facilmente si calcola

$$F_{\min(T_A,T_B)}(t) = 1 - (1 - F_{T_A}(t))(1 - F_{T_B}(t))$$

che è una legge esponenziale di parametro 2λ. Pertanto

$$\rho_T(t) = \int_0^t 2\lambda^2 \exp(-\lambda s)\exp(-2\lambda(t-s))\mathrm{d}s = 2\lambda(\exp(-\lambda t) - \exp(-2\lambda t))1\!\!1_{(0,+\infty)}(t)$$

da cui si ottiene la funzione di ripartizione

$$F_T(t) = \int_{-\infty}^t \rho_T(s)\mathrm{d}s = (1 + \exp(-2\lambda t) - 2\exp(-\lambda t))1\!\!1_{(0,+\infty)}(t).$$

Calcoliamo la media

$$E(T) = \int_{\mathbb{R}} t\rho_T(t)\mathrm{d}t = \frac{2}{\lambda} - \frac{1}{2\lambda} = \frac{3}{2\lambda} \equiv \frac{3}{2}E(T_C).$$

3.1) L'esercizio è del tutto analogo all'esercizio (2.2).

a)

$$F_Y(t) = \mathbb{P}(X^2 \leq t) = \begin{cases} 0 & x \leq 0 \\ \sqrt{t} & t \in (0,1) \\ 1 & t \geq 1 \end{cases}$$

$$\rho_Y(t) = \begin{cases} 0 & x \in (-\infty, \leq 0] \cup [1, +\infty) \\ \frac{1}{2\sqrt{t}} & t \in (0,1). \end{cases}$$

b) Si osservi che $\mathbb{P}(X = 0) = 0$ quindi con probabilità 1 l'inverso X^{-1} è ben definito (o, in linguaggio matematico, X^{-1} è *quasi certamente* definito). Per completezza, definiamo $Y := 0$ se $X = 0$.

$$F_Y(t) = \mathbb{P}(X^{-1} \le t) = \begin{cases} \mathbb{P}(X \in [1/t, 0)) = 0 & t < 0 \\ \mathbb{P}(X \le 0) = 0 & t = 0 \\ \mathbb{P}(X \ge 1/t) + \mathbb{P}(X \le 0) = 0 & t \in (0, 1] \\ \mathbb{P}(X \ge 1/t) + \mathbb{P}(X \le 0) = 1 - 1/t & t > 1 \end{cases}$$

$$\rho_Y(t) = \begin{cases} 0 & x \notin (1, +\infty) \\ \frac{1}{t^2} & t \in (1, +\infty). \end{cases}$$

c)

$$F_Y(t) = \mathbb{P}(\exp(X) \le t) = \begin{cases} 0 & x \le 0 \\ F_X(\ln(t)) = 0 & t \in (0, 1] \\ F_X(\ln(t)) = \ln(t) & t \in (1, e] \\ F_X(\ln(t)) = 1 & t > e \end{cases}$$

$$\rho_Y(t) = \begin{cases} 0 & x \notin (1, e] \\ \frac{1}{t} & t \in (1, e]. \end{cases}$$

3.2) Se $X_1, X_2, \dots, X_n$ sono indipendenti ed identicamente distribuite (con funzione di ripartizione F) e se $Y_1 \le Y_2 \le \cdots \le Y_n$ sono le stesse veriabili puntualmente ordinate, si ha che

$$F_{Y_k}(t) = \sum_{i=k}^{n} \binom{n}{i} F(t)^i (1 - F(t))^{k-i}$$

(provare!!!). Allora

$$F_{\min(X_1,\dots,X_n)}(t) = 1 - (1 - F(t))^n = \begin{cases} 0 & t \le 0 \\ t(2 - t) & t \in (0, 1] \\ 1 & t > 1 \end{cases}$$

$$F_{\max(X_1,\dots,X_n)}(t) = F(t)^n = \begin{cases} 0 & t \le 0 \\ t^2 & t \in (0, 1] \\ 1 & t > 1 \end{cases}$$

pertanto

$$\rho_{\min(X_1,\dots,X_n)(t)}(t) = 2(1 - t)\mathbb{1}_{(0,1)}(t)$$
$$\rho_{\max(X_1,\dots,X_n)(t)}(t) = 2t\mathbb{1}_{(0,1)}(t).$$

Quindi $E(\min(X_1, \dots, X_n)) = 1/3$, $E(\max(X_1, \dots, X_n)) = 2/3$ e $Var(\min(X_1, \dots, X_n)) = Var(\min(X_1, \dots, X_n)) = 1/18$.

3.3)

a)

$$\mathbb{P}(X \ge 1) = \int_1^{+\infty} \frac{1}{\pi(1 + x^2)} \mathrm{d}x = \frac{1}{\pi} \arctan(x)\Big|_1^{+\infty} = \frac{1}{4}.$$

b) Poichè la funzione densità è pari si ha che $\mathbb{P}(|X| \ge 1) = 2\mathbb{P}(X \ge 1) = 1/2$.