

# 1

## Qualche esempio..

### 1.1

#### 1.1.1

Sia  $\xi$  una variabile aleatoria avente densità  $f$ , di media 0 e varianza 1.

Determinare media e varianza della variabile aleatoria  $\eta = 3\xi + 7$ .

☞ ...

Si ha

$$\mathcal{P}(\xi \leq x) = \int_{-\infty}^x f(t) dt$$

e

$$\mathcal{P}(3\xi + 7 \leq x) = \mathcal{P}\left(\xi \leq \frac{x-7}{3}\right) = \int_{-\infty}^{\frac{x-7}{3}} f(t) dt = \int_{-\infty}^x \frac{1}{3} f\left(\frac{s-7}{3}\right) ds$$

Quindi

$$\phi(s) = \frac{1}{3} f\left(\frac{s-7}{3}\right)$$

è la PDF della variabile aleatoria  $\eta = 3\xi + 7$ .

Ne segue che la sua media  $\mu$  e la sua varianza  $\sigma$  possono essere calcolate mediante le

$$\begin{aligned} \mu &= \int_{-\infty}^x s \phi(s) ds = \int_{-\infty}^{+\infty} \frac{s}{3} f\left(\frac{s-7}{3}\right) ds = \\ &= \int_{-\infty}^{+\infty} (3t+7) f(t) dt = 3 \int_{-\infty}^{+\infty} t f(t) dt + 7 \int_{-\infty}^{+\infty} f(t) dt = 7 \end{aligned}$$

mentre

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{+\infty} (s-7)^2 \phi(s) ds = \int_{-\infty}^{+\infty} \frac{(s-7)^2}{3} f\left(\frac{s-7}{3}\right) ds = \\ &= \int_{-\infty}^{+\infty} 9s^2 f(s) ds = 9 \end{aligned}$$

e

$$\sigma = 3$$



## 1.1.2

Si considerino nel piano i punti

$$(0,4) (1,4) (2,3) (3,2) (4,2)$$

Determinare la retta di regressione ed il coefficiente di correlazione tra le variabili  $x$  ed  $y$

↔ ...

Chiamiamo  $x = (0, 1, 2, 3, 4)$  ed  $y = (4, 4, 3, 2, 2)$  i vettori delle ascisse e delle ordinate dei punti assegnati.

Sia

$$y = ax + b$$

la retta di regressione ad essi relativa.

Si ha che deve essere

$$\bar{y} = a\bar{x} + b$$

dove  $\bar{x} = \frac{1}{5} \sum_{i=1}^5 x_i$  e  $\bar{y} = \frac{1}{5} \sum_{i=1}^5 y_i$  sono le medie, rispettivamente dei valori assunti da  $x$  e da  $y$  e che

$$a = \frac{\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^5 (x_i - \bar{x})^2}$$

mentre il coefficiente di correlazione  $r$  è dato da

$$r = \frac{\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^5 (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^5 (y_i - \bar{y})^2}}$$

Eseguendo i calcoli si trova che

$$\bar{x} = 2$$

$$\bar{y} = 3$$

$$a = -0.6$$

$$b = 4.2$$

$$r = -0.9487$$

A titolo di esempio riportiamo le istruzioni mediante le quali possiamo eseguire i calcoli usando Matlab  $x=[0 \ 1 \ 2 \ 3 \ 4]$

```
y=[4 4 3 2 2]
polyfit(x,y,1)
xm=mean(x)
ym=mean(y)
a=(sum((x-xm).*(y-ym)))/(sum((x-xm).^2))
b=ym-a*xm
r=(sum((x-xm).*(y-ym)))/...
(sqrt(sum((x-xm).^2))*sqrt(sum((y-ym).^2)))
u=a*x+b
```

Possiamo anche trovare la retta di regressione nella forma

$$x = cy + d$$

In tal caso useremo formule simili alle precedenti dove il ruolo di  $x$  e di  $y$  è scambiato. I calcoli ci forniscono

$$c = -1.5$$

$$d = 6.5$$

e possono essere eseguiti con Matlab come segue

```
c=(sum((x-xm).*(y-ym)))/(sum((y-ym).^2))
d=xm-c*ym
```

$v=(x-d)/c$  I vettori  $u$  e  $v$  che sono stati calcolati, forniscono le ordinate delle rette di regressione che abbiamo trovato. Usando l'istruzione

```
plot(x,y,x,u,x,v)
```

possiamo disegnare i grafici delle due rette ed osservare come esse approssimino i valori dati.

Osserviamo inoltre che il valore del coefficiente di correlazione è vicino ad 1 e quindi i dati sono ben correlati, cioè le due rette che abbiamo trovato, che passano necessariamente per il punti  $(\bar{x}, \bar{y})$ , formano un angolo piccolo il cui coseno è proprio il coefficiente di correlazione  $r$ .

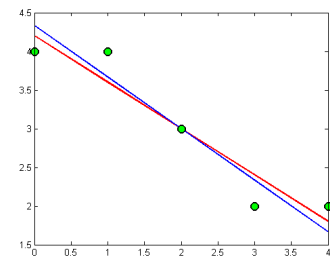


Figure 1.1:



### 1.1.3

Determinare la probabilità di ottenere, in 100 lanci di una moneta (non truccata), un numero di 'teste' compreso tra 45 e 60.

∞ ...

La variabile aleatoria  $b$  che restituisce il numero di successi (teste) ottenuti in  $n = 100$  ripetizioni di una prova bernoulliana (lancio della moneta) in cui la probabilità di successo è  $p = 0.5$ , è binomiale di media

$$\mu = np = 50$$

e varianza

$$\sigma^2 = npq = np(1 - p) = 25 \quad , \quad \sigma = 5$$

È noto che, per  $n$  grande ( $np > 5$ )  $\frac{b-\mu}{\sigma}$  si può approssimare con una variabile aleatoria gaussiana standardizzata  $\xi$  per cui:

$$\begin{aligned} \mathcal{P}(45 \leq b \leq 60) &= \mathcal{P}\left(\frac{45 - \mu}{\sigma} \leq \frac{b - \mu}{\sigma} \leq \frac{60 - \mu}{\sigma}\right) = \\ &= \mathcal{P}\left(\frac{45 - \mu}{\sigma} \leq \xi \leq \frac{60 - \mu}{\sigma}\right) = F\left(\frac{60 - \mu}{\sigma}\right) - F\left(\frac{45 - \mu}{\sigma}\right) \end{aligned}$$

dove  $F$  rappresenta come al solito la CDF della variabile Gaussiana standardizzata.

Possiamo calcolare che

$$F\left(\frac{60 - \mu}{\sigma}\right) - F\left(\frac{45 - \mu}{\sigma}\right) \approx 0.8186$$

usando Matlab, mediante l'istruzione

`normcdf((60-50)/5,0,1)-normcdf((45-50)/5,0,1)`

ovviamente lo stesso risultato si ottiene mediante la

`normcdf(60,50,5)-normcdf(45,50,5)`

possiamo anche apportare una correzione per tenere conto del fatto che  $b$  è discreta mentre  $\xi$  è continua e calcolare:

$$F\left(\frac{60.5 - \mu}{\sigma}\right) - F\left(\frac{44.5 - \mu}{\sigma}\right) \approx 0.8465$$

mediante la `normcdf((60.5-50)/5,0,1)-normcdf((44.5-50)/5,0,1)`

Infine possiamo anche usare direttamente la distribuzione binomiale calcolando

$$\mathcal{P}(45 \leq b \leq 60) = G(b) - G(a) \approx 0.8468$$

dove  $G$  rappresenta la CDF della variabile binomiale di media 50 e varianza 5, mediante la

`data=[45:1:60]`

`pdata=binopdf(data,100,0.5)`

`dd=sum(pdata)`

∞

1.2

1.2.1

Si considerino nel piano i punti

$$(0,0) (1,0) (1,1) (2,1)$$

Determinare la retta di regressione tra le variabili  $x$  ed  $y$  ...

Chiamiamo  $x = (0,1,1,2)$  ed  $y = (0,0,1,1)$  i vettori delle ascisse e delle ordinate dei punti assegnati.

Sia

$$y = ax + b$$

la retta di regressione ad essi relativa.

Si ha che deve essere

$$\bar{y} = a\bar{x} + b$$

dove  $\bar{x} = \frac{1}{4} \sum_{i=1}^4 x_i$  e  $\bar{y} = \frac{1}{4} \sum_{i=1}^4 y_i$  sono le medie, rispettivamente dei valori assunti da  $x$  e da  $y$  e che

$$a = \frac{\sum_{i=1}^4 (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^4 (x_i - \bar{x})^2}$$

mentre il coefficiente di correlazione  $r$  è dato da

$$r = \frac{\sum_{i=1}^4 (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^4 (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^4 (y_i - \bar{y})^2}}$$

Eseguendo i calcoli si trova che

$$\bar{x} = 1$$

$$\bar{y} = \frac{1}{2}$$

$$a = \frac{1}{2}$$

$$b = 0$$

$$r = 0.7071$$

A titolo di esempio riportiamo le istruzioni mediante le quali possiamo eseguire i calcoli usando Matlab `x=[0 1 1 2]`

`y=[0 0 1 1]`

`polyfit(x,y,1)`

`xm=mean(x)`

```

ym=mean(y)
a=(sum((x-xm).*(y-ym))/(sum((x-xm).^2))
b=ym-a*xm
r=(sum((x-xm).*(y-ym))/...
(sqrt(sum((x-xm).^2))*sqrt(sum((y-ym).^2)))
u=a*x+b

```

Possiamo anche trovare la retta di regressione nella forma

$$x = cy + d$$

Useremo formule simili alle precedenti dove il ruolo di  $x$  e di  $y$ , è scambiato. I calcoli ci forniscono

$$c = 1$$

$$d = \frac{1}{2}$$

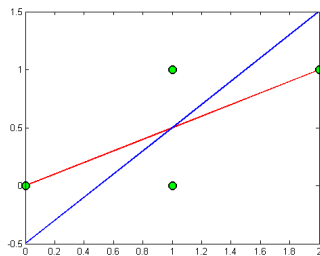


Figure 1.2:

e possono essere eseguiti con Matlab come segue

```

c=(sum((x-xm).*(y-ym))/(sum((y-ym).^2))
d=xm-c*ym
v=(x-d)/c

```

I vettori  $u$  e  $v$  che sono stati calcolati, forniscono le ordinate delle rette di regressione che abbiamo trovato. Usando l'istruzione `plot(x,y,x,u,x,v)` possiamo disegnare i grafici delle due rette ed osservare come esse approssimino i valori dati.

Il valore del coefficiente di correlazione è abbastanza vicino ad 1 e quindi i dati sono abbastanza ben correlati: le due rette che abbiamo trovato, che passano necessariamente per il punti  $(\bar{x}, \bar{y})$ , formano un angolo abbastanza piccolo il cui coseno è il coefficiente di correlazione  $r$ .

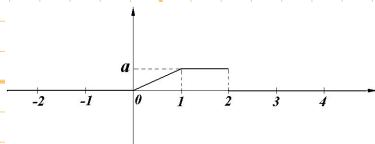


Figure 1.3:

### 1.2.2

Sia  $f$ , la funzione rappresentata nel seguente grafico

Determinarne  $a$  in modo che  $f$  rappresenti la densità di una variabile aleatoria  $X$  e successivamente calcolarne media e varianza.

...

Affinchè  $f$  rappresenti la densità di probabilità di una variabile aleatoria occorre che sia positiva e quindi dovrà intanto essere  $a > 0$ .

Inoltre deve aversi

$$\begin{aligned} 1 &= \int_{-\infty}^{+\infty} f(x) dx = \int_0^1 ax dx + \int_1^2 adx = \\ &= a \frac{x^2}{2} \Big|_0^1 + ax \Big|_1^2 = \frac{a}{2} + 2a - a = \frac{3}{2}a \end{aligned}$$

Pertanto  $a = \frac{2}{3}$  e per calcolare media e varianza possiamo procedere come segue.

$$\begin{aligned} \mu &= \int_{-\infty}^{+\infty} xf(x) dx = \int_0^1 ax^2 dx + \int_1^2 ax dx = a \frac{x^3}{3} \Big|_0^1 + a \frac{x^2}{2} \Big|_1^2 = \\ &= \frac{1}{3}a + \frac{4}{2}a - \frac{1}{2}a = \frac{11}{6}a \end{aligned}$$

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{+\infty} x^2 f(x) dx - \left(\frac{11}{6}a\right)^2 = \int_0^1 ax^3 dx + \int_1^2 ax^2 dx - \left(\frac{11}{6}a\right)^2 = \\ &= a \frac{x^4}{4} \Big|_0^1 + a \frac{x^3}{3} \Big|_1^2 - \left(\frac{11}{6}a\right)^2 = \frac{1}{4}a + \frac{8}{3}a - \frac{1}{3}a - \left(\frac{11}{6}a\right)^2 = \\ &= \frac{31}{12}a - \left(\frac{11}{6}a\right)^2 = \frac{74}{324} \end{aligned}$$



### 1.2.3

Si consideri un dado sulle cui 6 facce sono segnati i punteggi 1, 1, 2, 2, 3, 3 (tutti con eguale probabilità); determinare la probabilità che la somma dei punteggi ottenuti in 150 lanci sia compresa tra 280 e 310.

☞ ...

La variabile aleatoria  $b$  che restituisce la somma dei punteggi si può ottenere come somma di  $n = 150$  variabili aleatorie  $x_i$  la cui media e varianza sono date da

$$\mu_i = \frac{1}{3}(1+2+3) = 2 \quad , \quad \sigma_i^2 = \frac{1}{3}((1-2)^2 + (2-2)^2 + (3-2)^2) = \frac{2}{3}$$

Pertanto la media e la varianza di  $b$  sono date da

$$\mu = n\mu_i = 300 \quad , \quad \sigma^2 = n\sigma_i^2 \quad , \quad \sigma = \sqrt{n}\sigma_i = \sqrt{150 \frac{2}{3}} = 10$$

Per il teorema limite centrale la variabile aleatoria  $\frac{b-\mu}{\sigma}$  si può approssimare con una variabile aleatoria gaussiana standardizzata  $\xi$  per cui:

$$\begin{aligned}\mathcal{P}(280 \leq b \leq 310) &= \mathcal{P}\left(\frac{280-\mu}{\sigma} \leq \frac{b-\mu}{\sigma} \leq \frac{310-\mu}{\sigma}\right) = \\ &= \mathcal{P}\left(\frac{280-\mu}{\sigma} \leq \xi \leq \frac{310-\mu}{\sigma}\right) = F\left(\frac{310-\mu}{\sigma}\right) - F\left(\frac{280-\mu}{\sigma}\right)\end{aligned}$$

dove  $F$  rappresenta la CDF della variabile Gaussiana standardizzata.

Possiamo calcolare che

$$F\left(\frac{310-\mu}{\sigma}\right) - F\left(\frac{280-\mu}{\sigma}\right) \approx 0.8186$$

Possiamo anche apportare una correzione per tenere conto del fatto che  $b$  è discreta mentre  $\xi$  è continua e calcolare:

$$F\left(\frac{310.5-\mu}{\sigma}\right) - F\left(\frac{279.5-\mu}{\sigma}\right) \approx 0.8460$$

I risultati riportati si possono ottenere dalle seguenti istruzioni Matlab

```
dado=[1,1,2,2,3,3]
```

```
dadom=mean(dado)
```

```
dadov=var(dado,1)
```

```
mu=150*dadom
```

```
sigma=sqrt(150*dadov)
```

```
normcdf(310,mu,sigma)-normcdf(280,mu,sigma)
```

```
normcdf((310-mu)/sigma,0,1)-normcdf((280-mu)/sigma,0,1)
```

```
normcdf((310.5-mu)/sigma,0,1)-normcdf((279.5-mu)/sigma,0,1)
```

Vale la pena osservare che

```
normcdf(310,mu,sigma)-normcdf(280,mu,sigma)
```

```
normcdf((310-mu)/sigma,0,1)-normcdf((280-mu)/sigma,0,1)
```

eseguono lo stesso calcolo e che si è calcolato la varianza del vettore `dado` mediante l'istruzione `var(dado,1)`.

Matlab infatti, dato un vettore  $x = (x_i)$  di dati equiprobabili calcola

$$\text{var}(x) = \frac{1}{n-1} \sum_1^n x_i = \text{var}(x,0) \quad \text{e} \quad \text{var}(x,1) = \frac{1}{n} \sum_1^n x_i$$



### 1.3

#### 1.3.1

Si considerino nel piano i punti

$$(1, 1/2) \quad (2, 5/2) \quad (3, 7/2) \quad (4, 7/2)$$



Determinare la retta di regressione tra le variabili  $x$  ed  $y$  ed il coefficiente di correlazione.

☞ ...

Chiamiamo  $x = (1, 2, 3, 4)$  ed  $y = (1/2, 5/2, 7/2, 7/2)$  i vettori delle ascisse e delle ordinate dei punti assegnati.

Sia

$$y = ax + b$$

la retta di regressione ad essi relativa.

Deve essere

$$\bar{y} = a\bar{x} + b$$

dove  $\bar{x} = \frac{1}{4} \sum_{i=1}^4 x_i$  e  $\bar{y} = \frac{1}{4} \sum_{i=1}^4 y_i$  sono le medie, rispettivamente dei valori di  $x$  e da  $y$  e che

$$a = \frac{\sum_{i=1}^4 (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^4 (x_i - \bar{x})^2}$$

mentre il coefficiente di correlazione  $r$  è dato da

$$r = \frac{\sum_{i=1}^4 (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^4 (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^4 (y_i - \bar{y})^2}}$$

Eseguendo i calcoli si trova che

$$\bar{x} = 2.5$$

$$\bar{y} = 2.5$$

$$a = 1$$

$$b = 0$$

$$r = 0.9129$$

A titolo di esempio riportiamo le istruzioni mediante le quali possiamo eseguire i calcoli usando Matlab `x=[1 2 3 4]`

```
y=[1/2 5/2 7/2 7/2]
```

```
polyfit(x,y,1)
```

```
xm=mean(x)
```

```
ym=mean(y)
```

```
a=(sum((x-xm).*(y-ym)))/(sum((x-xm).^2))
```

```
b=ym-a*xm
```

```
r=(sum((x-xm).*(y-ym)))/...
```

```
(sqrt(sum((x-xm).^2))*sqrt(sum((y-ym).^2)))
```

```
u=a*x+b
```

Possiamo anche trovare la retta di regressione nella forma

$$x = cy + d$$

Useremo formule simili alle precedenti dove il ruolo di  $x$  e di  $y$  è scambiato. I calcoli ci forniscono

$$c = \frac{5}{6}$$

$$d = \frac{5}{12}$$

e possono essere eseguiti con Matlab come segue

$$c = (\text{sum}((x - xm) .* (y - ym)) / (\text{sum}((y - ym) .^ 2)))$$

$$d = xm - c * ym$$

$v = (x - d) / c$  I vettori  $u$  e  $v$  che sono stati calcolati, forniscono le ordinate delle rette di regressione che abbiamo trovato. Usando l'istruzione

$$\text{plot}(x, y, x, u, x, v)$$

possiamo disegnare i grafici delle due rette ed osservare come esse approssimino i valori dati.

Il valore del coefficiente di correlazione è abbastanza vicino ad 1 e quindi i dati sono abbastanza ben correlati: le due rette che abbiamo trovato, che passano necessariamente per il punti  $(\bar{x}, \bar{y})$ , formano un angolo abbastanza piccolo il cui coseno è il coefficiente di correlazione  $r$ .

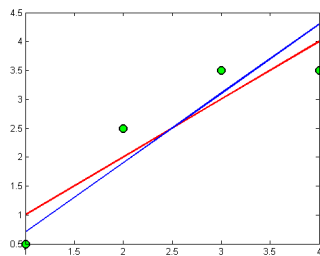


Figure 1.4:



### 1.3.2

Sia  $\zeta$ , la variabile aleatoria la cui funzione densità è definita da

$$\varphi(x) = \begin{cases} 2x & x \in [0, 1] \\ 0 & \text{altrove} \end{cases}$$

Calcolare media e varianza di  $\zeta$ .

...

La media è data da

$$\mu = \int_{-\infty}^{+\infty} x\varphi(x)dx = \int_0^1 2x^2 dx = 2 \frac{x^3}{3} \Big|_0^1 = \frac{2}{3}$$

mentre la varianza è

$$\begin{aligned}\sigma^2 &= \int_{-\infty}^{+\infty} x^2 \varphi(x) dx - \left(\frac{2}{3}\right)^2 = \int_0^1 2x^3 dx - \frac{4}{9} = 2 \frac{x^4}{4} \Big|_0^1 - \frac{4}{9} = \\ &= \frac{2}{4} - \frac{4}{9} = \frac{1}{18}\end{aligned}$$



1.4

1.4.1

Si considerino nel piano i punti

$$(0,0) (1,1) (1,3) (2,4)$$

Determinare la retta di regressione tra le variabili  $x$  ed  $y$  ed il coefficiente di correlazione.

∞ ...

Chiamiamo  $x = (0, 1, 1, 2)$  ed  $y = (0, 1, 3, 4)$  i vettori delle ascisse e delle ordinate dei punti assegnati.

Sia

$$y = ax + b$$

la retta di regressione ad essi relativa.

Deve essere

$$\bar{y} = a\bar{x} + b$$

dove  $\bar{x} = \frac{1}{4} \sum_{i=1}^4 x_i$  e  $\bar{y} = \frac{1}{4} \sum_{i=1}^4 y_i$  sono le medie, rispettivamente dei valori di  $x$  e da  $y$  e che

$$a = \frac{\sum_{i=1}^4 (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^4 (x_i - \bar{x})^2}$$

mentre il coefficiente di correlazione  $r$  è dato da

$$r = \frac{\sum_{i=1}^4 (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^4 (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^4 (y_i - \bar{y})^2}}$$

Eseguendo i calcoli si trova che

$$\begin{aligned}\bar{x} &= 1 \\ \bar{y} &= 2 \\ a &= 2 \\ b &= 0 \\ r &= 0.8944\end{aligned}$$

A titolo di esempio riportiamo le istruzioni mediante le quali possiamo eseguire i calcoli usando Matlab `x=[0 1 1 2]`

```
y=[0 1 3 4]
polyfit(x,y,1)
xm=mean(x)
ym=mean(y)
a=(sum((x-xm).*(y-ym)))/(sum((x-xm).^2))
b=ym-a*xm
r=(sum((x-xm).*(y-ym)))/...
(sqrt(sum((x-xm).^2))*sqrt(sum((y-ym).^2)))
u=a*x+b
```

Possiamo anche trovare la retta di regressione nella forma

$$x = cy + d$$

Useremo formule simili alle precedenti dove il ruolo di  $x$  e di  $y$  è scambiato. I calcoli ci forniscono

$$\begin{aligned}c &= \frac{2}{5} \\ d &= \frac{1}{5}\end{aligned}$$

e possono essere eseguiti con Matlab come segue

```
c=(sum((x-xm).*(y-ym)))/(sum((y-ym).^2))
d=xm-c*ym
v=(x-d)/c
```

I vettori  $u$  e  $v$  che sono stati calcolati, forniscono le ordinate delle rette di regressione che abbiamo trovato. Usando l'istruzione

```
plot(x,y,x,u,x,v)
```

possiamo disegnare i grafici delle due rette ed osservare come esse approssimino i valori dati.

Il valore del coefficiente di correlazione è abbastanza vicino ad 1 e quindi i dati sono abbastanza ben correlati: le due rette che abbiamo trovato, che passano necessariamente per il punto  $(\bar{x}, \bar{y})$ , formano un angolo abbastanza piccolo il cui coseno è il coefficiente di correlazione  $r$ .

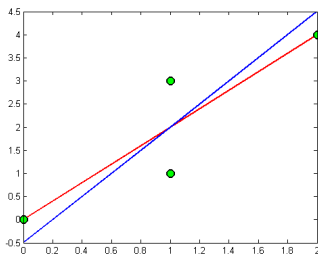


Figure 1.5:

## 1.4.2

	B	N
I	7	3
II	5	5
III	3	7

Si considerino tre urne *I*, *II* e *III* che contengono Gettoni *Bianchi* e *Neri* come riportato nella tabella a fianco. Si sceglie un'urna casualmente e si estrae un gettone dall'urna prescelta.

La probabilità di uscita di ciascuna urna è  $\frac{1}{3}$ . Calcolare la probabilità che l'estrazione sia avvenuta dall'urna *I*, dall'urna *II* o dall'urna *III* sapendo che è stato estratto un gettone *Nero*.

🔗 ...

Indichiamo con

$$\mathcal{P}(C|U)$$

la probabilità di estrarre un Gettone di colore *C* dall'urna *U*. (Probabilità di estrarre un gettone di colore *C* condizionata al fatto che l'estrazione avvenga dall'urna *U*)

Avremo che

$$\begin{aligned} \mathcal{P}(N|I) &= \frac{3}{10} & \mathcal{P}(B|I) &= \frac{7}{10} \\ \mathcal{P}(N|II) &= \frac{1}{2} & \mathcal{P}(B|II) &= \frac{1}{2} \\ \mathcal{P}(N|III) &= \frac{7}{10} & \mathcal{P}(B|III) &= \frac{3}{10} \end{aligned}$$

Inoltre, poichè ognuna delle tre urne può essere scelta con egual probabilità avremo

$$\mathcal{P}(I) = \mathcal{P}(II) = \mathcal{P}(III)$$

e, per il teorema di Bayes

$$\begin{aligned} \mathcal{P}(I|N) &= \frac{\mathcal{P}(N|I)\mathcal{P}(I)}{\mathcal{P}(N|I)\mathcal{P}(I) + \mathcal{P}(N|II)\mathcal{P}(II) + \mathcal{P}(N|III)\mathcal{P}(III)} = \\ &= \frac{\frac{3}{10} \frac{1}{3}}{\frac{3}{10} \frac{1}{3} + \frac{1}{2} \frac{1}{3} + \frac{7}{10} \frac{1}{3}} = \frac{3}{15} \\ \mathcal{P}(II|N) &= \frac{\mathcal{P}(N|II)\mathcal{P}(II)}{\mathcal{P}(N|I)\mathcal{P}(I) + \mathcal{P}(N|II)\mathcal{P}(II) + \mathcal{P}(N|III)\mathcal{P}(III)} = \\ &= \frac{\frac{5}{10} \frac{1}{3}}{\frac{3}{10} \frac{1}{3} + \frac{1}{2} \frac{1}{3} + \frac{7}{10} \frac{1}{3}} = \frac{5}{15} \\ \mathcal{P}(III|N) &= \frac{\mathcal{P}(N|III)\mathcal{P}(III)}{\mathcal{P}(N|I)\mathcal{P}(I) + \mathcal{P}(N|II)\mathcal{P}(II) + \mathcal{P}(N|III)\mathcal{P}(III)} = \\ &= \frac{\frac{7}{10} \frac{1}{3}}{\frac{3}{10} \frac{1}{3} + \frac{1}{2} \frac{1}{3} + \frac{7}{10} \frac{1}{3}} = \frac{7}{15} \end{aligned}$$



Si consideri una moneta truccata in modo che *Testa* esca con probabilità  $\frac{1}{5}$  e *Croce* esca con probabilità  $\frac{4}{5}$ . Determinare la probabilità che si abbia un numero di teste compreso tra 1980 e 2040, avendo lanciato la moneta 10000 volte. ...

Sia  $b$  la variabile aleatoria che restituisce il numero di successi (teste) ottenuti in  $n = 10000$  ripetizioni della prova bernoulliana effettuata lanciando una moneta con probabilità di ottenere Testa  $p = 0.2$  e probabilità di ottenere croce  $q = 0.8$ .

$b$  è binomiale di media

$$\mu = np = 2000$$

e varianza

$$\sigma^2 = npq = 1600, \quad \sigma = 40$$

È noto che, per  $n$  grande ( $np > 5$ )  $\frac{b-\mu}{\sigma}$  si può approssimare con una variabile aleatoria gaussiana standardizzata  $\xi$  per cui:

$$\begin{aligned} \mathcal{P}(1980 \leq b \leq 2040) &= \mathcal{P}\left(\frac{1980 - \mu}{\sigma} \leq \frac{b - \mu}{\sigma} \leq \frac{2040 - \mu}{\sigma}\right) = \\ &= \mathcal{P}\left(\frac{1980 - \mu}{\sigma} \leq \xi \leq \frac{2040 - \mu}{\sigma}\right) = F\left(\frac{2040 - \mu}{\sigma}\right) - F\left(\frac{1980 - \mu}{\sigma}\right) \end{aligned}$$

dove  $F$  è la CDF della variabile Gaussiana standardizzata.

Possiamo calcolare che

$$F\left(\frac{2040 - \mu}{\sigma}\right) - F\left(\frac{1980 - \mu}{\sigma}\right) \approx 0.532807$$

usando Matlab, mediante l'istruzione

```
normcdf((2040-2000)/40,0,1)-normcdf((1980-2000)/40,0,1)
```

ovviamente lo stesso risultato si ottiene mediante la

```
normcdf(2040,2000,40)-normcdf(1980,2000,40)
```

possiamo anche apportare una correzione per tenere conto del fatto che  $b$  è discreta mentre  $\xi$  è continua e calcolare:

$$F\left(\frac{60.5 - \mu}{\sigma}\right) - F\left(\frac{44.5 - \mu}{\sigma}\right) \approx 0.5402$$

mediante la

```
normcdf((2040.5-2000)/40,0,1)-normcdf((1979.5-2000)/40,0,1)
```

Infine possiamo anche usare direttamente la distribuzione binomiale calcolando

$$\mathcal{P}(1980 \leq b \leq 2040) = G(b) - G(a) \approx 0.539541$$

dove  $G$  rappresenta la CDF della variabile binomiale di media 50 e varianza 5, mediante la

```
data=[1980:1:2040]
pdata=binopdf(data,10000,0.2)
dd.tex=sum(pdata)
```



1.5

Si consideri una variabile aleatoria  $x$  la cui distribuzione di probabilità (PDF) è del tipo

$$f(t) = \begin{cases} 2t - t^2 & t \in [0, 1] \\ \frac{a}{t^4} & t > 1 \\ 0 & t < 0 \end{cases}$$

Determinare  $a$  in modo che  $f$  sia una distribuzione di probabilità

...

Affinchè  $f$  sia una distribuzione di probabilità occorre innanzi tutto che  $a \geq 0$ . Inoltre dovrà essere

$$\int_{-\infty}^{+\infty} f(t) dt = 1$$

Quindi

$$\begin{aligned} 1 &= \int_{-\infty}^{+\infty} f(t) dt = \int_0^1 2t - t^2 dt + \int_1^{+\infty} \frac{a}{t^4} dt = \\ &= t^2 - \frac{1}{3}t^3 \Big|_0^1 - a \frac{t^{-3}}{3} \Big|_1^{+\infty} = 1 - \frac{1}{3} + a \frac{1}{3} = 1 - (1-a) \frac{1}{3} \end{aligned}$$

e si dovrà scegliere  $a = 1$ .

Calcolare la media  $\mu$  e la varianza  $\sigma^2$  di  $x$

∞ ...

$$\begin{aligned} \mu &= \int_{-\infty}^{+\infty} t f(t) dt = \int_0^1 2t^2 - t^3 dt + \int_1^{+\infty} \frac{1}{t^3} dt = \\ &= \frac{2}{3}t^3 - \frac{1}{4}t^4 \Big|_0^1 - \frac{t^{-2}}{2} \Big|_1^{+\infty} = \frac{2}{3} - \frac{1}{4} + \frac{1}{2} = \frac{11}{12} \end{aligned}$$

inoltre

$$\sigma^2 = \int_{-\infty}^{+\infty} (t - \mu)^2 f(t) dt = \left( \int_{-\infty}^{+\infty} t^2 f(t) dt \right) - \mu^2$$

e

$$\begin{aligned} \int_{-\infty}^{+\infty} t^2 f(t) dt &= \int_0^1 2t^3 - t^4 dt + \int_1^{+\infty} \frac{1}{t^2} dt = \\ &= \frac{2}{4}t^4 - \frac{1}{5}t^5 \Big|_0^1 - \frac{1}{t} \Big|_1^{+\infty} = \frac{2}{4} - \frac{1}{5} + 1 = \frac{13}{10} \end{aligned}$$

Da cui

$$\sigma^2 = \int_{-\infty}^{+\infty} (t - \mu)^2 f(t) dt = \frac{13}{10} - \left( \frac{11}{12} \right)^2 \approx 1.2236$$

$$\sigma \approx$$

∞ ×



Calcolare la probabilità che la variabile aleatoria  $x$  assuma valori compresi tra 0 e 4  $\Leftrightarrow$  ...

$$\begin{aligned} \mathcal{P}(0 \leq x \leq 4) &= \int_0^4 t^2 f(t) dt = \int_0^1 2t - t^2 dt + \int_1^4 \frac{1}{t^4} dt = \\ &= t^2 - \frac{1}{3}t^3 \Big|_0^1 - \frac{t^{-3}}{3} \Big|_1^4 = 1 - \frac{1}{3} + \frac{1}{3} - \frac{1}{192} = \frac{191}{192} \approx 0.9948 \end{aligned}$$

$\Leftrightarrow$  ✘

Stimare usando la disuguaglianza di Tchebichev

$$\mathcal{P}(|x - \mu| > 1)$$

$\Leftrightarrow$  ...

$$\mathcal{P}(|x - \mu| > 1) \leq \sigma^2 \approx 1.2236$$

*Evidentemente in questo caso la disuguaglianza di Tchebichev non produce risultati significativi.*

$\Leftrightarrow$  ✘

1.5.1

Per andare dal punto  $A$  al punto  $B$  posso seguire la strada  $I$  o la strada  $II$ .

Nel 80% dei casi scelgo la strada  $II$ .

La probabilità di avere ritardo seguendo la strada  $I$  è del 10%

La probabilità di avere ritardo seguendo la strada  $II$  è del 15%

Poichè sono arrivato in ritardo qual è la probabilità che io abbia seguito la strada  $I$

$\Leftrightarrow$  ...

Indichiamo con  $\mathcal{P}(I)$ , ( $\mathcal{P}(II)$ ), la probabilità di scegliere la strada I, (II), con  $\mathcal{P}(R)$  la probabilità di avere ritardo e con  $\mathcal{P}(R')$  la probabilità di non avere ritardo.

$$\mathcal{P}(I) = 0.2 \quad , \quad \mathcal{P}(II) = 0.8$$

$$\mathcal{P}(R|I) = 0.1 \quad , \quad \mathcal{P}(R|II) = 0.15$$

Quindi per la regola di Bayes,

$$\mathcal{P}(I|R) = \frac{\mathcal{P}(R|I)\mathcal{P}(I)}{\mathcal{P}(R|I)\mathcal{P}(I) + \mathcal{P}(R|II)\mathcal{P}(II)} = \frac{0.1 * 0.2}{0.1 * 0.2 + 0.15 * 0.8} = \frac{1}{7}$$



Se non sono arrivato in ritardo qual è la probabilità che io abbia seguito la strada I

... ..

$$\mathcal{P}(R'|I) = 0.9 \quad , \quad \mathcal{P}(R'|II) = 0.85$$

Quindi per la regola di Bayes,

$$\mathcal{P}(I|R') = \frac{\mathcal{P}(R'|I)\mathcal{P}(I)}{\mathcal{P}(R'|I)\mathcal{P}(I) + \mathcal{P}(R'|II)\mathcal{P}(II)} = \frac{0.9 * 0.2}{0.9 * 0.2 + 0.85 * 0.8} = \frac{9}{43}$$



## 1.6

Si consideri una variabile aleatoria  $\zeta$  la cui distribuzione di probabilità (PDF) è del tipo

$$f(t) = \begin{cases} e^{at} & t \leq 0 \\ e^{bt} & t > 0 \end{cases}$$

Determinare  $a, b$  in modo che  $f$  sia una distribuzione di probabilità e che  $\zeta$  abbia media 0.5. ... ..

Poichè  $f$  è una funzione positiva, affinché sia una PDF, sarà sufficiente verificare che

$$\int_{-\infty}^{+\infty} f(t) dt = 1$$

Dovrà pertanto aversi che  $a > 0$  e  $b < 0$  perchè sia garantita la convergenza dell'integrale che definisce  $\mu$  ed inoltre poichè

$$\begin{aligned} \int_{-\infty}^{+\infty} f(t) dt &= \int_{-\infty}^0 f(t) dt + \int_0^{+\infty} f(t) dt = \\ &= \int_{-\infty}^0 e^{at} dt + \int_0^{+\infty} e^{bt} dt = \frac{1}{a} e^{at} \Big|_{-\infty}^0 + \frac{1}{b} e^{bt} \Big|_0^{+\infty} = \\ &= \frac{1}{a} - \frac{1}{b} = \frac{b-a}{ab} \end{aligned}$$

dovrà risultare

$$\frac{b-a}{ab} = 1$$

e

$$ab = b - a$$

La media della variabile aleatoria definita da  $f$  è data da

$$\mu = \int_{-\infty}^{+\infty} tf(t) dt$$

e

$$\begin{aligned} \int_{-\infty}^{+\infty} tf(t) dt &= \int_{-\infty}^0 te^{at} dt + \int_0^{+\infty} te^{bt} dt = \\ &= \frac{e^{at}}{a} t \Big|_{-\infty}^0 - \int_{-\infty}^0 \frac{e^{at}}{a} dt + \frac{e^{bt}}{b} t \Big|_0^{+\infty} - \int_0^{+\infty} \frac{e^{bt}}{b} dt = \\ &= -\frac{1}{a^2} + \frac{1}{b^2} \end{aligned}$$

Per cui dovrà aversi

$$\frac{1}{b^2} - \frac{1}{a^2} = \frac{1}{2}$$

Pertanto per soddisfare le richieste occorre che

$$\begin{cases} 1 = \frac{b-a}{ab} \\ \frac{1}{2} = \frac{a^2-b^2}{a^2b^2} = \frac{a-b}{ab} \frac{a+b}{ab} = -\frac{a+b}{ab} \end{cases}$$

e

$$b - a = ab = -2(a + b)$$

Se ne deduce che

$$\begin{cases} a = -3b \\ 4b = -3b^2 \end{cases}$$

da cui

$$a = 4 \quad b = -\frac{4}{3}$$



Stabilire se è più probabile che  $\xi$  assuma valori maggiori o minori della media. ...

Si ha

$$\mathcal{P}(\xi \geq \mu) = \int_{\frac{1}{2}}^{+\infty} e^{-\frac{4}{3}t} dt = -\frac{3}{4} e^{-\frac{4}{3}t} \Big|_{\frac{1}{2}}^{+\infty} = \frac{3}{4} e^{-\frac{4}{3} \cdot \frac{1}{2}} = \frac{3}{4} e^{-\frac{2}{3}} \approx 0.3851$$

Ovviamente

$$\mathcal{P}(\xi \leq \mu) = 1 - \mathcal{P}(\xi \geq \mu)$$

come possiamo anche calcolare direttamente

$$\begin{aligned} \mathcal{P}(\xi \leq \mu) &= \int_{-\infty}^0 e^{4t} dt + \int_0^{\frac{1}{2}} e^{-\frac{4}{3}t} dt = \frac{1}{4} e^{4t} \Big|_{-\infty}^0 - \frac{3}{4} e^{-\frac{4}{3}t} \Big|_0^{\frac{1}{2}} = \\ &= \frac{1}{4} + \frac{3}{4} - \frac{3}{4} e^{-\frac{2}{3}} = 1 - \frac{3}{4} e^{-\frac{2}{3}} \approx \end{aligned}$$

Quindi

$$\mathcal{P}(\xi \leq \mu) > \mathcal{P}(\xi \geq \mu)$$



Calcolare la probabilità che la variabile aleatoria  $x$  assuma valori compresi tra  $-1$  e  $1$

...

$$\begin{aligned} \mathcal{P}(-1 \leq \xi \leq 1) &= \int_{-1}^0 e^{4t} dt + \int_0^1 e^{-\frac{4}{3}t} dt = \\ &= \int_{-1}^0 e^{4t} dt + \int_0^1 e^{-\frac{4}{3}t} dt = \frac{1}{4} e^{4t} \Big|_{-1}^0 - \frac{3}{4} e^{-\frac{4}{3}t} \Big|_0^1 = \\ &= \frac{1}{4} (1 - e^{-4}) + \frac{3}{4} (1 - e^{-\frac{4}{3}}) \approx \end{aligned}$$



Calcolare la probabilità che la variabile aleatoria  $\xi$  assuma valori compresi tra  $-5$  e  $1$  condizionata al fatto che  $\xi$  assuma valori positivi.

$$\begin{aligned} \mathcal{P}(-5 \leq \xi \leq 1 | \xi \geq 0) &= \frac{\mathcal{P}(0 \leq \xi \leq 1)}{\mathcal{P}(\xi > 0)} = \\ &= \frac{\int_0^1 e^{-\frac{4}{3}t} dt}{\int_0^{+\infty} e^{-\frac{4}{3}t} dt} = \frac{-\frac{3}{4}e^{-\frac{4}{3}t} \Big|_0^1}{-\frac{3}{4}e^{-\frac{4}{3}t} \Big|_0^{+\infty}} = \frac{\frac{3}{4}(1 - e^{-\frac{4}{3}})}{\frac{3}{4}} = 1 - e^{-\frac{4}{3}} \approx \end{aligned}$$

Sia  $\xi$  lo scarto in grammi dal valore nominale del peso di una confezione di merce.

Esaminando una particolare confezione si rileva che pesa 1.5 grammi più del peso nominale; se si rigetta l'ipotesi che  $\xi$  abbia come PDF la funzione  $f$  determinata in precedenza, calcolare la probabilità di commettere un errore di I specie (rigettare l'ipotesi nel caso l'ipotesi sia vera)

...

*Nel caso in cui  $\xi$  rappresenti lo scarto in grammi dal valore nominale, la probabilità di avere una confezione con peso superiore a quello nominale di 1.5 grammi è*

$$\mathcal{P}(\xi > 1.5) = \int_{\frac{3}{2}}^{+\infty} e^{-\frac{4}{3}t} dt = -\frac{3}{4}e^{-\frac{4}{3}t} \Big|_{\frac{3}{2}}^{+\infty} = \frac{3}{4}e^{-\frac{4}{3} \cdot \frac{3}{2}} = \frac{3}{4}e^{-2} \approx 0.1015$$

*Quindi se, avendo osservato un peso superiore a quello nominale di 1.5 grammi, rigettiamo l'ipotesi che  $\xi$  rappresenti lo scarto, commetteremo un errore con probabilità del 10%.*

*Ciò significa che, con un livello di confidenza  $\alpha = 90\%$ , avendo osservato un peso superiore a quello nominale di 1.5 grammi, rigetteremo l'ipotesi che  $\xi$  rappresenti lo scarto.*



## 1.7

Un'azienda dispone di due linee di produzione per tubi di diametro (medio)  $d = 100\text{mm}$ . La prima linea (Linea1) produce tubi il cui diametro è distribuito normalmente con scarto quadratico  $\sigma_1 = 1\text{mm}$  mentre la seconda (Linea2) produce tubi il cui diametro è distribuito normalmente con scarto quadratico  $\sigma_2 = 0.5\text{mm}$ .

La produzione media giornaliera è di 300 pezzi di cui 180 prodotti dalla prima linea e 120 prodotti dalla seconda. Calcolare la probabilità che un pezzo prodotto dalla Linea1 abbia diametro compreso tra  $99.5\text{mm}$  e  $100.5\text{mm}$ .

↻ ...

Sia  $\delta_i$  la variabile aleatoria che rappresenta il diametro dei tubi prodotti dalla Linea  $i$ ;  $\delta_i$  è una variabile gaussiana di media  $d$  e varianza  $\sigma_i$ .

Sia

$$\xi_i = \frac{\delta_i - d}{\sigma_i}$$

$\xi_1$  è una variabile aleatoria gaussiana standardizzata e si ha

$$\begin{aligned} \mathcal{P}(99.5 \leq \delta_1 \leq 100.5) &= \mathcal{P}\left(\frac{99.5 - d}{\sigma_1} \leq \frac{\delta_1 - d}{\sigma_1} \leq \frac{100.5 - d}{\sigma_1}\right) = \\ &= \mathcal{P}(-0.5 \leq \xi_1 \leq 0.5) = F(0.5) - F(-0.5) \approx 0.691 - 0.309 \approx 0.382 \end{aligned}$$

dove  $F$  indica come al solito la densità di probabilità cumulativa della variabile normale standardizzata. Calcolando con Excel

$$=\text{Normdist.tex}(0.5,0,1,\text{TRUE}) - \text{Normdist.tex}(-0.5,0,1,\text{TRUE}) \approx 0.3829.$$

↻ ✘

Calcolare la probabilità che un pezzo prodotto dalla Linea1 abbia diametro minore di  $99.5\text{mm}$  e maggiore di  $100.5\text{mm}$ . ↻ ...

Similmente

$$\begin{aligned} \mathcal{P}(\delta_1 \leq 99.5) &= \mathcal{P}\left(\frac{\delta_1 - d}{\sigma_1} \leq \frac{99.5 - d}{\sigma_1}\right) = \\ &= \mathcal{P}(\xi_1 \leq -0.5) = F(-0.5) \approx 0.309 \end{aligned}$$

e

$$\begin{aligned}\mathcal{P}(\delta_1 \geq 100.5) &= \mathcal{P}\left(\frac{\delta_1 - d}{\sigma_1} \geq \frac{100.5 - d}{\sigma_1}\right) = \\ &= \mathcal{P}(\xi_1 \geq 0.5) = 1 - F(0.5) \approx 0.309\end{aligned}$$

Calcolando con Excel

`Normdist.tex(-0.5,0,1,TRUE)` e `1-Normdist.tex(0.5,0,1,TRUE)`.

Calcolare la probabilità che un pezzo prodotto dalla Lineaz abbia diametro compreso tra 99.5mm e 100.5mm.

$$\begin{aligned}\mathcal{P}(99.5 \leq \delta_2 \leq 100.5) &= \mathcal{P}\left(\frac{99.5 - d}{\sigma_2} \leq \frac{\delta_2 - d}{\sigma_2} \leq \frac{100.5 - d}{\sigma_2}\right) = \\ &= \mathcal{P}(-1 \leq \xi_1 \leq 1) = F(1) - F(-1) \approx 0.841 - 0.159 \approx 0.683\end{aligned}$$

dove  $F$  indica come al solito la densità di probabilità cumulativa della variabile normale standardizzata.



Calcolare la probabilità che un pezzo prodotto dalla Lineaz abbia diametro minore di 99.5mm e maggiore di 100.5mm.

 ...

$$\mathcal{P}(\delta_2 \leq 99.5) = \mathcal{P}\left(\frac{\delta_2 - d}{\sigma_2} \leq \frac{99.5 - d}{\sigma_2}\right) = \mathcal{P}(\xi_2 \leq -1) = F(-0.5) \approx 0.159$$

e

$$\mathcal{P}(\delta_2 \geq 100.5) = \mathcal{P}\left(\frac{\delta_2 - d}{\sigma_2} \geq \frac{100.5 - d}{\sigma_2}\right) = \mathcal{P}(\xi_2 \geq 1) = 1 - F(1) \approx 0.159$$



Calcolare la probabilità che un pezzo prodotto abbia diametro minore di 99.5mm e maggiore di 100.5mm.

 ...

La probabilità che un tubo sia prodotto dalla linea  $i$  si può calcolare mediante le

$$\mathcal{P}(L_1) = \frac{180}{300} = \frac{3}{5} = 0.6 \quad \mathcal{P}(L_2) = \frac{120}{300} = \frac{2}{5} = 0.4$$

per cui, se  $\delta$  è la variabile aleatoria che rappresenta il diametro dei tubi prodotti

$$\begin{aligned} \mathcal{P}(\delta \leq 99.55) &= \mathcal{P}(\delta \leq 99.5|L_1)\mathcal{P}(L_1) + \mathcal{P}(\delta \leq 99.5|L_2)\mathcal{P}(L_2) = \\ &= \mathcal{P}(\delta_1 \leq 99.5)\mathcal{P}(L_1) + \mathcal{P}(\delta_2 \leq 99.5)\mathcal{P}(L_2) \approx \\ &\approx 0.309 * 0.6 + 0.159 * 0.4 \approx 0.249 \end{aligned}$$

$$\begin{aligned} \mathcal{P}(\delta \geq 100.55) &= \mathcal{P}(\delta \geq 100.5|L_1)\mathcal{P}(L_1) + \mathcal{P}(\delta \geq 100.5|L_2)\mathcal{P}(L_2) = \\ &= \mathcal{P}(\delta_2 \geq 100.5)\mathcal{P}(L_1) + \mathcal{P}(\delta_2 \geq 100.5)\mathcal{P}(L_2) \approx \\ &\approx 0.309 * 0.6 + 0.159 * 0.4 \approx 0.249 \end{aligned}$$

❖ ❖

Calcolare la probabilità che un pezzo prodotto abbia diametro compreso tra 99.5mm e 100.5mm. ❖ ...

*Similmente*

$$\begin{aligned} \mathcal{P}(99.5 \leq \delta \leq 100.5) &= \mathcal{P}(99.5 \leq \delta \leq 100.5|L_1)\mathcal{P}(L_1) + \mathcal{P}(99.5 \leq \delta \leq 100.5|L_2)\mathcal{P}(L_2) \\ &= \mathcal{P}(99.5 \leq \delta_1 \leq 100.5)\mathcal{P}(L_1) + \mathcal{P}(99.5 \leq \delta_2 \leq 100.5)\mathcal{P}(L_2) \approx 0.383 * 0.6 + 0.683 * 0.4 \approx \end{aligned}$$

❖ ❖

Vengono considerati difettosi i tubi il cui diametro differisce dalla media per più del 0.5%.

Calcolare la probabilità che un pezzo difettoso della produzione giornaliera provenga dalla Linea1 ❖ ...

Indichiamo con  $D$  l'evento il tubo è difettoso.



$$\mathcal{P}(D) = \mathcal{P}(\delta \geq 100.5.5) + \mathcal{P}(\delta \leq 99.5.5) = 0.498$$

$$\mathcal{P}(D|L_1) = \mathcal{P}(\delta_1 \geq 100.5.5) + \mathcal{P}(\delta_1 \leq 99.5.5) = 0.318$$

$$\mathcal{P}(D|L_2) = \mathcal{P}(\delta_2 \geq 100.5.5) + \mathcal{P}(\delta_2 \leq 99.5.5) = 0.618$$

Usando la Regola di Bayes otteniamo

$$\mathcal{P}(L_i|D) = \frac{\mathcal{P}(D|L_i)\mathcal{P}(L_i)}{\mathcal{P}(D|L_2)\mathcal{P}(L_2) + \mathcal{P}(D|L_1)\mathcal{P}(L_1)}$$

per cui

$$\mathcal{P}(L_1|D) = \frac{0.618 * 0.6}{0.618 * 0.6 + 0.318 * 0.4} \approx 0.744$$

$$\mathcal{P}(L_2|D) = \frac{0.318 * 0.4}{0.618 * 0.6 + 0.318 * 0.4} \approx 0.256$$



## 1.8

In un impianto le sferette prodotte vengono raccolte in due contenitori ai quali giungono dopo aver percorso un tubo che ad un certo punto si sdoppia in due rami che portano ciascuno ad uno dei due contenitori.

Le sferette hanno probabilità 0.5 di percorrere uno dei due rami e quindi di essere smistate nel corrispondente contenitore. Si calcoli quante sferette deve poter contenere ogni contenitore affinché nessuno dei due contenitori debba essere svuotato durante il giorno con un livello di significatività del 2.5%, tenendo conto che la produzione giornaliera è di 10000 sferette.



Una sferetta finisce nel contenitore  $C_1$  casualmente con probabilità  $p = 0.5$  e nel contenitore  $C_2$  probabilità  $q = 0.5$  e la prova è ripetuta per ognuna delle sferette prodotte. Quindi il numero delle sferette che finiscono nel contenitore  $C_i$  è descritto da una variabile aleatoria binomiale  $b_i$  relativa ai parametri

- Numero delle prove  $n = 10000$
- Probabilità di successo  $p = 0.5$  (o equivalentemente  $q = 0.5$ )

per cui si ha che media e varianza sono date da

$$\mu = np = 5000 \quad \sigma^2 = npq = 2500 \quad \sigma = 50$$

Se  $N$  è il numero di sferette che possono essere contenute in  $C_i$ , esso dovrà essere svuotato non appena si abbia che  $b_i > N$ ; se vogliamo che il contenitore non debba essere svuotato con un livello di significatività del 2.5% dovremo assicurarci che

$$\mathcal{P}(b_i > N) < 0.025$$

$b_i$ , essendo  $np > 5$ , può essere normalizzata ed approssimata mediante una variabile aleatoria gaussiana standard  $\xi_i$  e si ha che

$$\mathcal{P}(b_i > N) = \mathcal{P}\left(\frac{b_i - \mu}{\sigma} > \frac{N - \mu}{\sigma}\right) = \mathcal{P}\left(\xi_i > \frac{N - \mu}{\sigma}\right)$$

Dovremo quindi determinare  $N$  in modo che

$$0.025 > \mathcal{P}\left(\xi_i > \frac{N - \mu}{\sigma}\right) = 1 - \mathcal{P}\left(\xi_i < \frac{N - \mu}{\sigma}\right) = 1 - F\left(\frac{N - \mu}{\sigma}\right)$$

dove  $F$  indica come al solito la densità di probabilità cumulativa della variabile normale standardizzata.

Ne viene che

$$F\left(\frac{N - \mu}{\sigma}\right) > 1 - 0.025 = 0.975$$

e

$$\frac{N - \mu}{\sigma} > F^{-1}(0.975) \approx 1.96$$

come si vede ad esempio usando la seguente istruzione di Excel


`NormInv(0.975,0,1)`

Se ne conclude che deve essere

$$N > \mu + \sigma \cdot 1.96 = 5000 + 50 \cdot 1.96 = 5097.99 \approx 5098$$



Un dado viene lanciato 100 volte con l'intento di stabilire se è stato truccato in modo da alterare la probabilità di uscita dei numeri pari rispetto a quella dei numeri dispari.

Determinare i limiti entro i quali devono mantenersi i numeri di uscite di un pari affinché si possa sostenere che il dado non è truccato con un livello di significatività del 5%.  ...

L'uscita di un pari o di un dispari nel lancio di un dado è il risultato di

una prova bernoulliana in cui probabilità di successo e di insuccesso valgono  $p = q = 1/2$ . Il numero di pari o di dispari può quindi essere rappresentato da una variabile aleatoria  $\delta$  binomiale con  $n = 100$  di media e varianza date da

$$\mu = np = 100 * 0.5 = 50 \quad \sigma^2 = npq = 100 * 0.25 = 25 \quad \sigma = 5$$

Possiamo ritenere che il dado non sia truccato se il numero di pari o di dispari non è troppo lontano dal valore medio  $\mu$ . Cioè se

$$\mu - k \leq \delta \leq \mu + k$$

Naturalmente nulla impedisce che un dado non truccato restituisca su 100 lanci 100 numeri pari, tuttavia la probabilità che ciò accada è piccola e quindi diremo che siamo certi che il dado non è truccato con un livello di significatività del 5% se

$$\mathcal{P}((\delta < \mu - k) \cup (\delta > \mu + k)) < 0.05$$

ovvero se

$$\mathcal{P}((\delta - \mu < -k) \cup (\delta - \mu > k)) < 0.05$$

Dal momento che  $np = 50$  possiamo usare una variabile aleatoria gaussiana standardizzata  $\xi$  per approssimare ognuna delle  $\delta - \mu$ ,  $\pi - \mu$  dopo che siano state standardizzate ed avremo che

$$\begin{aligned} 0.05 &= \mathcal{P}((\delta - \mu < -k) \cup (\delta - \mu > k)) = \\ &= \mathcal{P}\left(\left(\frac{\delta - \mu}{\sigma} < -\frac{k}{\sigma}\right) \cup \left(\frac{\delta - \mu}{\sigma} > \frac{k}{\sigma}\right)\right) = \\ &= \mathcal{P}\left(\left(\xi < -\frac{k}{\sigma}\right) \cup \left(\xi > \frac{k}{\sigma}\right)\right) = 2 * \left(1 - F\left(\frac{k}{\sigma}\right)\right) \end{aligned}$$

dove  $F$  indica la densità di probabilità cumulativa della variabile normale standardizzata.

Ne segue che

$$2 * \left(1 - F\left(\frac{k}{\sigma}\right)\right) = 0.05$$

e

$$\frac{k}{\sigma} > F^{-1}(0.975) \approx 1.96$$

Se ne conclude che deve essere

$$k > \sigma * 1.96 = 5 * 1.96 = 9.8 = 9$$

per cui riterremo non truccato il dado se il numero di pari, o equivalentemente di dispari, uscito è compreso tra 41 e 59.



Determinare i limiti entro i quali devono mantenersi i numeri di uscite di un pari affinché si possa sostenere che il dado non è truccato in modo da favorire l'uscita di un pari con un livello di significatività del 5%.

↻ ...

*Nel caso in cui ci interessi stabilire che il dado non favorisce l'uscita di pari dovremo determinare  $k$  in modo che*

$$\mathcal{P}(\delta > \mu + k) = 0.05$$

*Avremo, con le stesse notazioni precedenti,*

$$\mathcal{P}(\delta - \mu > k) = \mathcal{P}\left(\frac{\delta - \mu}{\sigma} > \frac{k}{\sigma}\right) = 1 - F\left(\frac{k}{\sigma}\right) = p(\xi) > \frac{k}{\sigma} = 0.05$$

e

$$\frac{k}{\sigma} = F^{-1}(0.95) = 1.645$$

*(con Excel `NormInv(0.95,0,1)`) da cui*

$$k = \sigma 1.645 = 5 * 1.645 = 8.2 \approx 8$$

*per cui riterremo non truccato il dado se il numero di dispari, uscito inferiore a 58.*



1.9

Un'azienda produce stagionalmente 300 cassette ciascuna contenente 24 mele. La produzione è suddivisa in due classi di qualità A e B in proporzione del 30% e del 70%, rispettivamente, per la stagione in esame.

Supponendo di non effettuare nessun controllo prima di suddividere la produzione in cassette, calcolare la probabilità che una cassetta contenga 0, 1 o al più una mela di categoria B. ↻ ...

*La produzione stagionale ammonta a*

$$300 \times 24 = 7200 \quad \text{mele}$$

Indichiamo con  $m \in A$  e  $m \in B$  il fatto che una mela  $m$  sia di classe  $A$  o  $B$ , rispettivamente, avremo che

$$\mathcal{P}(m \in A) = \frac{3}{10} = 0.3 = p \quad , \quad \mathcal{P}(m \in B) = \frac{7}{10} = 0.7 = q$$

Poichè la produzione è stagionale e le cassette sono confezionate durante l'arco di tutta la stagione supponiamo che la probabilità  $\mathcal{P}(m \in A)$  e  $\mathcal{P}(m \in B)$  rimanga costante.

Lo riempimento di una cassetta è un processo bernoulliano e se intendiamo contare le mele di classe  $B$  che essa contiene definiamo 'successo' il fatto di scegliere una mela di classe  $B$ . Sia  $\xi$  la variabile aleatoria che restituisce il numero di mele di classe  $B$  contenute nella cassetta.

$$\mathcal{P}(\xi = 0) = \binom{0}{24} q^0 p^{24} = \left(\frac{3}{10}\right)^{24} \approx 0.28 \times 10^{-12}$$

$$\mathcal{P}(\xi = 1) = \binom{1}{24} q^1 p^{23} = 24 \left(\frac{7}{10}\right) \left(\frac{3}{10}\right)^{23} \approx 0.16 \times 10^{-10}$$

mentre

$$\mathcal{P}(\xi \leq 1) = \mathcal{P}(\xi = 0) + \mathcal{P}(\xi = 1) \approx 0.16 \times 10^{-10}$$



Si supponga poi di introdurre un controllo di qualità in grado di riconoscere che una mela è di classe  $A$  con certezza e di attribuire ad una mela di classe  $B$  la qualifica di classe  $A$  nel 5% dei casi.

Calcolare la probabilità che una cassetta di classe  $A$  contenga 0, 1 o al più una mela di categoria  $B$  nell'ipotesi che venga effettuato tale controllo.

↪ ...

Dopo aver introdotto il controllo di qualità le mele sono divise in due categorie  $C_A$  e  $C_B$  e vengono confezionate le cassette considerando di classe  $A$  le mele in  $C_A$ .

$C_A$  conterrà le mele effettivamente di classe  $A$  più quelle di classe  $B$  classificate di classe  $A$  dal controllo di qualità. quindi in tutto

$$(300 \times 24) \left( \frac{3}{10} + \frac{7}{10} \times \frac{5}{100} \right) = (300 \times 24) \left( \frac{13}{40} \right) = 2340$$

mele che serviranno per riempire

$$(300) \left( \frac{13}{40} \right) \approx 97$$

cassette che saranno definite di classe A nel confezionare le quali si procederà estraendo dalla classe A in cui però sarà possibile trovare mele di classe B con probabilità

$$p_C = \mathcal{P}(B|C_A) = \frac{\mathcal{P}(C_A|B)\mathcal{P}(B)}{\mathcal{P}(C_A|B)\mathcal{P}(B) + \mathcal{P}(C_A|A)\mathcal{P}(A)} =$$

$$= \frac{\frac{5}{100} \frac{7}{10}}{\frac{5}{100} \frac{7}{10} + 1 \frac{3}{10}} = \frac{7}{67} \approx 0.1$$

La probabilità che una cassetta selezionata di classe A contenga 0, 1 o al più una mela di classe B è data da una variabile bernoulliana  $\xi_C$  relativa ad una probabilità di successo  $p_C$  e si ha


$$\mathcal{P}(\xi_C = 0) = \binom{0}{24} q_C^0 p_C^{24} = \left( \frac{7}{67} \right)^{24} \approx 10^{-24}$$

$$\mathcal{P}(\xi_C = 1) = \binom{1}{24} q_C^1 p_C^{23} = 24 \left( \frac{60}{67} \right) \left( \frac{7}{67} \right)^{23} \approx 10^{-23}$$

mentre

$$\mathcal{P}(\xi_C \leq 1) = \mathcal{P}(\xi_C = 0) + \mathcal{P}(\xi_C = 1) \approx 10^{-23}$$



In questa seconda ipotesi quante cassette di classe A vengono confezionate?  ...

Vengono confezionate 97 cassette

Supponendo che una cassetta non selezionata possa essere venduta ad un prezzo  $p$ , che una cassetta di presunta classe A possa subire un incremento di prezzo del 30% e una cassetta di classe B una diminuzione del 20%, calcolare quanto è conveniente investire nel controllo di qualità.

↻ ...

nel caso non si introduca il controllo di qualità si potranno vendere 300 cassette a prezzo  $p$  ricavando

$$R = 300p$$

, mentre se si introduce il controllo si potranno vendere 97 cassette a prezzo  $p + p\frac{3}{10}$  e 203 cassette a prezzo  $p - p\frac{2}{10}$  con un ricavo

$$R_C = 97 \left( p + p\frac{3}{10} \right) + 203 \left( p - p\frac{2}{10} \right) = p * \frac{577}{2}$$

Il confronto tra  $R$  ed  $R_C$  mostra che non è conveniente introdurre il controllo. ↻ ✖

Osserviamo che l'uso della distribuzione binomiale non è corretto in quanto il processo di scelta delle mele per lo riempimento delle cassette è un processo di estrazione senza reintroduzione.

Per tenere conto di questo fatto occorrerebbe usare la distribuzione ipergeometrica in luogo della distribuzione binomiale. tuttavia poiché il numero di mele è grande possiamo approssimare la distribuzione ipergeometrica con la distribuzione binomiale senza perdere significatività.

Se infatti supponiamo che  $\xi$  sia ipergeometrica di parametri 5040 (casi favorevoli), 2160 (casi contrari) e 24 (numero delle estrazioni), possiamo calcolare che

$$\mathcal{P}(\xi = 0) \approx 0.26 \times 10^{-12}$$

$$\mathcal{P}(\xi = 1) \approx 0.15 \times 10^{-10}$$

mentre

$$\mathcal{P}(\xi \leq 1) = \mathcal{P}(\xi = 0) + \mathcal{P}(\xi = 1) \approx 0.15 \times 10^{-10}$$

Come è evidente i risultati sono praticamente identici. Riportiamo le istruzioni per calcolare con MAPLE le probabilità cercate:

Ad esempio

```
> statevalf[pf,binomiald[24,.7]](1);
```

per calcolare il valore della distribuzione di probabilità binomiale relativa a 24 estrazioni con probabilità di successo .7 in 1. (probabilità di ottenere 1 successo in 24 tentativi con probabilità di successo 0.7).

```
> statevalf[pf,hypergeometric[5040, 2160, 24]](1);
```

per calcolare il valore della distribuzione di probabilità ipergeometrica di parametri 5040 (casi favorevoli), 2160 ( casi contrari) e 24 ( numero delle estrazioni) in 1. (probabilità di ottenere 1 successo estraendo per 24 volte da una popolazione in cui ci sono 5040 casi favorevoli e 2160 casi contrari).

Concludiamo poi riportando i grafici e le istruzioni per ottenerli con MAPLE delle due distribuzioni di probabilità usate.

```
> with(plots);
> with(stats):
> with(plottools);
> bino:=statevalf[pf,binomiald[24,.7]](t):
> bi :=unapply(bino,t):
> biGph:=plot([seq([[k-1,bi(k)],[k,bi(k)],[k,bi(k+1)]]
> ,k=1..23)],color=yellow, thickness=6):
> hype:=statevalf[pf,hypergeometric[5040, 2160, 24]](t):
> hy :=unapply(hype,t):
> hyGph:=plot([seq([[k-1,hy(k)],[k,hy(k)],[k,hy(k+1)]]
> ,k=1..23)],color=blue, thickness=1):
> display(hyGph,biGph);
```

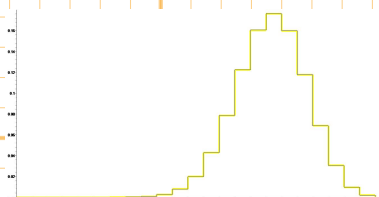


Figure 1.6:



## 1.10

Si consideri una prova bernoulliana in cui la probabilità di successo  $p$  è incognita. Si consideri poi la variabile aleatoria  $\xi$  che restituisce il numero di successi ottenuti su  $n$  prove effettuate. Esprimere la media  $\mu = \mu(n, p)$  e lo scarto quadratico medio  $\sigma = \sigma(n, p)$  di  $\xi$  in funzione del numero di ripetizioni  $n$  e della probabilità di successo  $p$ .

Si osservano i valori assunti da  $\xi$  in  $m$  casi e si rileva che  $\xi$  assume valori maggiori di  $k$  in  $a$  casi si chiede, alternativamente, di stimare  $p$  in funzione di  $n, m, k, a$  oppure per  $n = 100, m = 10000, k = 75, a = 8732$ .

↻ ...

La variabile aleatoria  $\xi$  ha una distribuzione binomiale la cui media e varianza sono

$$\mu(n, p) = np \quad , \quad \sigma^2(n, p) = np(1 - p)$$

Per  $np > 5$  possiamo standardizzare  $\xi$  ed approssimarla mediante una gaussiana di media 0 e varianza 1.

Se consideriamo un campione costituito da  $m$  valori di cui  $a$  hanno valore superiore a  $k$  avremo che, detta  $F$  la CDF Gaussiana standard,

$$1 - F\left(\frac{k - \mu}{\sigma}\right) = \left(\frac{\xi - \mu}{\sigma} > \frac{k - \mu}{\sigma}\right) = \frac{a}{m}$$

da cui

$$F\left(\frac{k - \mu}{\sigma}\right) = 1 - \frac{a}{m}$$

Ne segue che deve risultare

$$\frac{k - \mu}{\sigma} = F^{-1}\left(1 - \frac{a}{m}\right) = \alpha$$

e

$$(k - np) = \alpha \sqrt{np(1 - p)}$$

da cui

$$(k - np)^2 = \alpha^2 np(1 - p)$$

Svolgendo i calcoli si ottiene un'equazione di secondo grado

$$p^2(n^2 + \alpha^2 n) - p(2kn + \alpha^2 n) + k^2 = 0$$

che ha per soluzioni

$$\frac{2kn + \alpha^2 n \pm \sqrt{4n^2 k^2 \alpha^2 + \alpha^4 n^2 - 4k^2 \alpha^2 n}}{2n(n + \alpha^2)}$$

Delle due soluzioni va scelta quella in corrispondenza della quale  $k - np$  ed  $\alpha$  hanno lo stesso segno in quanto elevando al quadrato abbiamo perso tale informazione.

Di seguito è riportato il codice Maple per eseguire i calcoli nell'esempio numerico proposto.

```
> with(stats);
> n:=100:
> m:=10000:
> m := 100000:
> k:=75:
> a:=8732:
> alpha:=statevalf[icdf,normald](1-a/m);
      alpha := 1.357444553
> eq:=p^2*(n^2+alpha^2*n)-p*(2*k*n+alpha^2*n) +k^2;
      eq := 10184.26557 p^2 - 15184.26557 p + 5625
> sol:=solve(eq,p);
      sol := 0.8038969807, 0.6870564382
> k-n*sol[1];
      -5.38969807
> k-n*sol[2];
      6.29435618
```

Dal momento che  $\alpha > 0$  avremo quindi che  $p = 0.68705$  in quanto in corrispondenza di tale soluzione si ha  $k = np > 0$ .

### 1.11

L'azienda **A** produce componenti elettronici che l'azienda **B** utilizza per la produzione delle sue apparecchiature. **B** quindi acquista da **A** componenti in partite di  $N$  pezzi ciascuna.

Ciascuna partita di componenti contiene pezzi difettosi con probabilità  $p_0$ , tuttavia, introducendo un controllo di qualità più efficiente, la probabilità di trovare pezzi difettosi si riduce a  $p_1$ .

**A** afferma di aver introdotto tale controllo ( $Cq$ ) e **B**, per verificare la veridicità dell'affermazione, effettua un campionamento estraendo, con reintroduzione,  $n$  pezzi e verificando se sono difettosi. **B** trova che degli  $n$  pezzi estratti  $q$  sono difettosi; indichiamo questo evento con  $D_q$ .

È vero che, nel caso in cui il controllo sia stato effettuato  $\frac{q}{n} = p_1$ ? Giustificare brevemente.

Calcolare la probabilità di  $D_q$  supponendo che  $\mathcal{P}(Cq) = 0$  cioè che il controllo di qualità non sia stato effettuato.

Calcolare la probabilità di  $D_q$  supponendo che  $\mathcal{P}(Cq) = 1$  cioè che il controllo di qualità sia stato effettuato.

Calcolare la probabilità di  $D_q$  supponendo assegnata la probabilità  $\mathcal{P}(Cq) = t$  che il controllo di qualità sia stato effettuato.

Se assumiamo che  $\mathcal{P}(Cq) = t$  come possiamo aggiornare la stima di  $\mathcal{P}(Cq)$  sapendo che  $D_q$  è accaduto?

☞ ...

Intanto non è chiaramente sempre vero che  $\frac{q}{n} = p_1$ ; infatti se ad esempio consideriamo una partita di  $N = 10000$  pezzi per i quali  $p_1 = 0.01$  e se ci capita di estrarre gli  $n = 100$  pezzi difettosi che tale partita contiene avremo  $q = 100$  e  $1 = \frac{q}{n} \neq p_1 = 0.01$ . Se  $\mathcal{P}(Cq) = 0$  avremo che la probabilità di estrarre un pezzo difettoso è  $p_0$  per cui

$$\mathcal{P}(D_q) = \binom{n}{q} p_0^q (1 - p_0)^{n-q}$$

trattandosi di un esperimento bernoulliano in cui la probabilità di successo (estrarre un pezzo difettoso) è  $p_0$ . Analogamente Se  $\mathcal{P}(Cq) = 1$  avremo che la probabilità di estrarre un pezzo difettoso è  $p_1$  per cui

$$\mathcal{P}(D_q) = \binom{n}{q} p_1^q (1 - p_1)^{n-q}$$

trattandosi di un esperimento bernoulliano in cui la probabilità di successo (estrarre un pezzo difettoso) è  $p_1$ .

Nel caso in cui sia noto che  $\mathcal{P}(Cq) = t$ , possiamo calcolare

$$\begin{aligned} \mathcal{P}(D_q) &= \mathcal{P}(D_q|Cq)\mathcal{P}(Cq) + \mathcal{P}(D_q|Cq^c)\mathcal{P}(Cq^c) = \\ &= \mathcal{P}(D_q|Cq)\mathcal{P}(Cq) + \mathcal{P}(D_q|Cq^c)(1 - \mathcal{P}(Cq)) \end{aligned}$$

e

$$\mathcal{P}(D_q) = t \left( \binom{n}{q} p_1^q (1 - p_1)^{n-q} \right) + (1 - t) \left( \binom{n}{q} p_0^q (1 - p_0)^{n-q} \right)$$

Possiamo aggiornare la stima di  $\mathcal{P}(Cq)$  sapendo che  $D_q$  è accaduto, calcolando

$$\begin{aligned} \mathcal{P}(Cq|D_q) &= \frac{\mathcal{P}(D_q|Cq)\mathcal{P}(Cq)}{\mathcal{P}(D_q|Cq)\mathcal{P}(Cq) + \mathcal{P}(D_q|Cq^c)\mathcal{P}(Cq^c)} = \\ &= \frac{t \binom{n}{q} p_1^q (1 - p_1)^{n-q}}{t \left( \binom{n}{q} p_1^q (1 - p_1)^{n-q} \right) + (1 - t) \left( \binom{n}{q} p_0^q (1 - p_0)^{n-q} \right)} \end{aligned}$$

Fin qui abbiamo supposto che il Controllo di qualità sia stato applicato all'intero lotto; nel caso in cui invece tale controllo sia stato applicato con

probabilità  $t$  ad ogni singolo pezzo, la probabilità  $\mathcal{P}(D_1)$  di estrarre un pezzo difettoso si può calcolare mediante la

$$\mathcal{P}(D_1) = \mathcal{P}(D_1|Cq)\mathcal{P}(Cq) + \mathcal{P}(D_1|Cq^c)\mathcal{P}(Cq^c) = p_1t + p_0(1-t)$$

e

$$\mathcal{P}(D'_q) = \binom{n}{q} (p_1t + p_0(1-t))^q (1 - p_1t - p_0(1-t))^{n-q} \neq \mathcal{P}(D'_q)$$

Ne deduciamo che gli eventi "estrazione di un pezzo" non sono, in tal caso, indipendenti.

Si ha che  $\mathcal{P}(D_q) = \mathcal{P}(D'_q)$  nel caso in cui si sappia che il controllo di qualità è effettuato con probabilità  $t$  su ogni singolo pezzo. In tal caso invece, gli eventi "estrazione di un pezzo" sono indipendenti.

### 1.12

Il numero di automobili in transito al casello di un località turistica durante il weekend dipende dalle condizioni meteorologiche e può essere descritto dalla variabile aleatoria  $x$  nel caso il tempo sia soleggiato  $y$  nel caso il tempo sia nuvoloso e  $z$  in caso di cattivo tempo.

Le densità di probabilità di  $x$ ,  $y$ , e  $z$  sono rispettivamente date da:

$x < 300$	$300 < x < 800$	$x > 800$
10%	30%	60%

$y < 300$	$300 < y < 800$	$y > 800$
20%	50%	30%

$z < 300$	$300 < z < 800$	$z > 800$
50%	40%	10%

Rappresentare graficamente la PDF di  $x$ ,  $y$ ,  $z$ . Supponendo che, durante l'estate, la probabilità di una giornata soleggiata sia del 65%, che si abbia un cielo nuvoloso nel 25% dei casi e che si abbia cattivo tempo nel restante 10%, determinare la PDF della variabile aleatoria che descrive il numero di auto in transito.

Considerato un week end in cui si sia registrato un numero di auto compreso tra 300 e 800, calcolare la probabilità che il tempo sia stato soleggiato.

Si consideri poi la seguente tabella che riporta i dati di 1000 osservazioni del numero di auto in transito durante l'inverno relativamente alle condizioni metereologiche

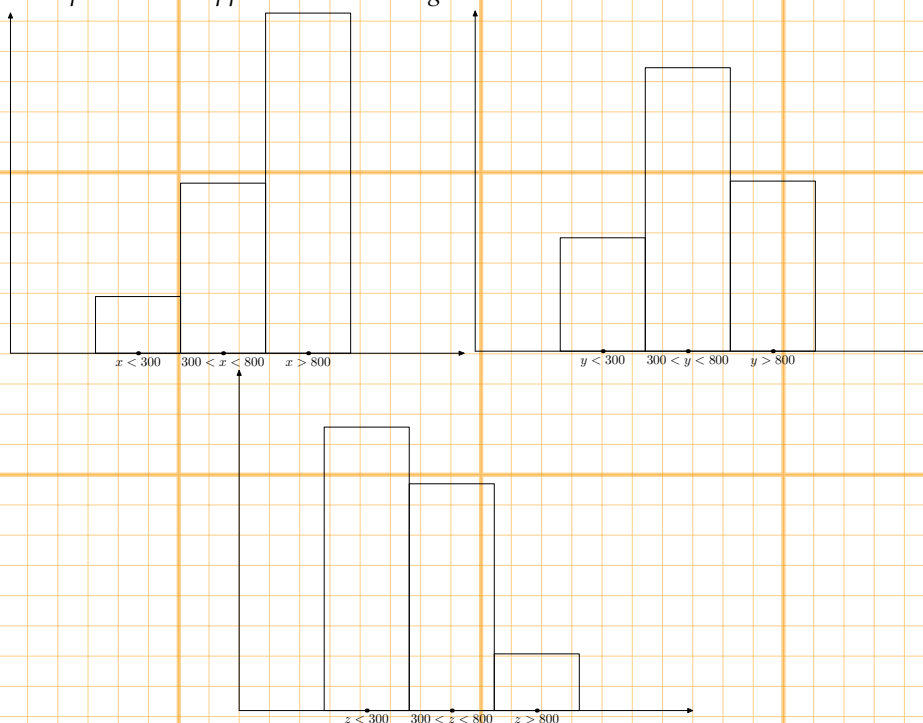
	$x < 300$	$300 < x < 800$	$x > 800$
Buone	100	300	150
Cattive	150	250	50

Determinare la distribuzione di probabilità delle variabili aleatorie  $\xi, \eta$  che rappresentano il numero di auto in transito rispettivamente quando ci sono buone o cattive condizioni meteorologiche. Determinare la distribuzione di probabilità delle variabili aleatorie  $\sigma_1, \sigma_2, \sigma_3$  che rappresentano lo stato delle condizioni meteorologiche nel caso in cui il numero di transiti sia, rispettivamente, inferiore a 300 compreso tra 300 e 800 o superiore a 800.

Stabilire se il numero di transiti è indipendente dalle condizioni meteorologiche.

☞ ...

Indichiamo con  $T_1$  l'evento il numero di transiti è inferiore a 300  $T_2$  l'evento il numero di transiti è compreso tra 300 ed 800  $T_3$  l'evento il numero di transiti è superiore a 800 La PDF delle variabili aleatorie  $x, y, z$  può essere rappresentata come segue:



inoltre, se indichiamo con  $S$  l'evento il tempo è soleggiato con  $N$  l'evento

il tempo è nuvoloso e con  $C$  l'evento il tempo è cattivo avremo

$$\mathcal{P}(S) = 0.65 \quad , \quad \mathcal{P}(N) = 0.25 \quad , \quad \mathcal{P}(C) = 0.10$$

e

$$\mathcal{P}(T_k) = \mathcal{P}(T_k|S)\mathcal{P}(S) + \mathcal{P}(T_k|N)\mathcal{P}(N) + \mathcal{P}(T_k|C)\mathcal{P}(C)$$

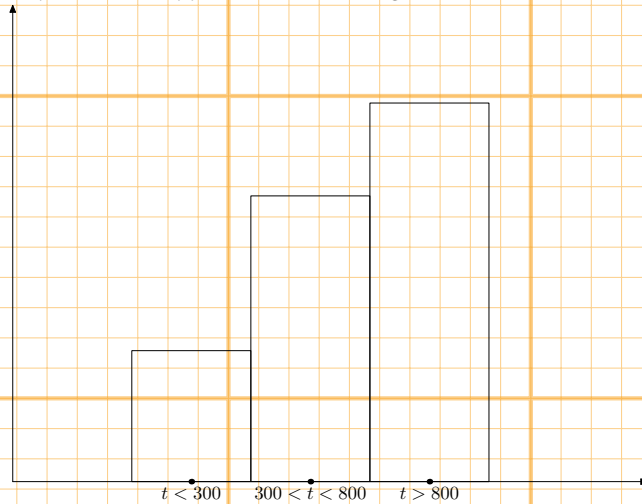
Si ottiene che

$$\mathcal{P}(T_1) = \frac{10}{100} \frac{65}{100} + \frac{20}{100} \frac{25}{100} + \frac{50}{100} \frac{10}{100} = \frac{650 + 500 + 500}{10000} = \frac{1650}{10000}$$

$$\mathcal{P}(T_2) = \frac{30}{100} \frac{65}{100} + \frac{50}{100} \frac{25}{100} + \frac{40}{100} \frac{10}{100} = \frac{1950 + 1250 + 400}{10000} = \frac{3600}{10000}$$

$$\mathcal{P}(T_3) = \frac{10}{100} \frac{65}{100} + \frac{20}{100} \frac{25}{100} + \frac{50}{100} \frac{10}{100} = \frac{650 + 500 + 500}{10000} = \frac{1650}{10000}$$

ed il grafico della PDF della variabile aleatoria  $t$  che rappresenta il numero dei transiti può essere rappresentata come segue.



Possiamo calcolare la probabilità di  $S$  ammesso che sia accaduto  $T_2$  come

$$\begin{aligned} \mathcal{P}(S|T_2) &= \frac{\mathcal{P}(T_2|S)\mathcal{P}(S)}{\mathcal{P}(T_2|S)\mathcal{P}(S) + \mathcal{P}(T_2|N)\mathcal{P}(N) + \mathcal{P}(T_2|C)\mathcal{P}(C)} = \\ &= \frac{30 \cdot 65}{30 \cdot 65 + 50 \cdot 25 + 40 \cdot 10} = \frac{1950}{1950 + 1250 + 400} = \frac{1950}{3600} = \frac{195}{360} \approx 0.5417 \end{aligned}$$

Manteniamo ora le notazioni per  $T_1$ ,  $T_2$  e  $T_3$  ed indichiamo con  $B$  l'evento il tempo è buono e con  $C$  l'evento il tempo è cattivo

Per quanto riguarda  $\xi$  e  $\eta$  si ha

$$\mathcal{P}(\xi < 300) = \frac{10}{55} \quad , \quad \mathcal{P}(300 < \xi < 800) = \frac{30}{55} \quad , \quad \mathcal{P}(\xi > 800) = \frac{15}{55}$$

$$\mathcal{P}(\eta < 300) = \frac{15}{45} \quad , \quad \mathcal{P}(300 < \eta < 800) = \frac{25}{45} \quad , \quad \mathcal{P}(\eta > 800) = \frac{5}{45}$$

mentre per identificare  $\sigma_1, \sigma_2, \sigma_3$  è sufficiente osservare che

$$\mathcal{P}(\sigma_1 = B) = \frac{10}{25}, \quad \mathcal{P}(\sigma_1 = C) = \frac{15}{25}$$

$$\mathcal{P}(\sigma_2 = B) = \frac{30}{55}, \quad \mathcal{P}(\sigma_2 = C) = \frac{25}{55}$$

$$\mathcal{P}(\sigma_3 = B) = \frac{15}{30}, \quad \mathcal{P}(\sigma_3 = C) = \frac{5}{30}$$

Per stabilire se il numero di transiti e le condizioni del tempo sono indipendenti possiamo osservare che la PDF delle due variabili aleatorie congiunte è la seguente.

	$x < 300$	$300 < x < 800$	$x > 800$	
Buone	0.1	0.3	0.15	0.55
Cattive	0.15	0.25	0.05	0.45
	0.25	0.55	0.20	1

Poichè, ad esempio, la probabilità che il tempo sia buono e che i transiti siano minori di 300 è 0.10 mentre la probabilità che il tempo sia buono è  $\frac{0.1}{0.25}$  mentre la probabilità che i transiti siano minori di 300 è  $\frac{0.1}{0.55}$  e quindi

$$\mathcal{P}(B \cap t < 300) = 0.1 \neq \frac{0.1}{0.25} \frac{0.1}{0.55} = \mathcal{P}(B)\mathcal{P}(t < 300)$$

