08 luglio 2016 es.1 (Gli esami a Matematica)

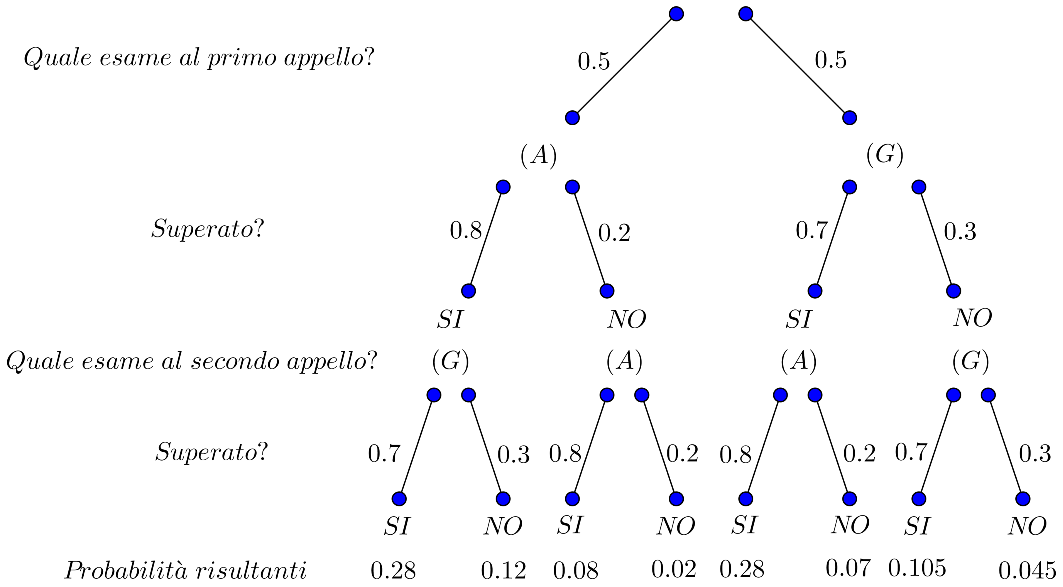
Paolo, studente del primo anno di Matematica, sta preparando gli esami di Analisi 1 e Geometria 1 per la sessione estiva, nella quale avrà due appelli per ciascuna materia. Paolo ritiene che sia 0.8 la probabilità di superare l'esame di Analisi, 0.7 di superare Geometria. Indeciso sul da farsi, tira a sorte per scegliere quale esame sostenere al primo appello; se lo supererà, al secondo appello affronterà l'altra materia, altrimenti ritenterà l'esame fallito (con invariata probabilità di superarlo). Si esclude che Paolo sostenga due esami nello stesso appello.

a) Qual è la probabilità che nella sessione estiva Paolo superi almeno un esame tra i due?

b) Impariamo che Paolo ha preso 24 in Analisi, al secondo appello. Qual è la probabilità che Paolo abbia superato anche Geometria nella sessione estiva?

Soluzione

a) Il seguente grafo illustra le diverse alternative che si possono presentare:



$$P(\text{almeno un esame superato}) = 1 - P(\text{nessun esame superato}) = 1 - (0.02 + 0.045) = 0.935$$

b) Siano A2 l'evento "Paolo supera Analisi al 2° appello", G1 l'evento "Paolo supera Geometria al 1° appello". Ci interessa $P(G \mid \text{A2})$. Questa è

$$P(\text{G1} \mid \text{A2}) = \frac{P(\text{G1} \wedge \text{A2})}{P(\text{A2})} = \frac{0.28}{0.08+0.28} = \frac{7}{9} \approx 0.778.$$

08 luglio 2016 es.2 (Le auto sportive di Lewis)

Lewis, appassionato di auto sportive, ne possiede 6: tre "Mercedes", due "Ferrari", una "Lamborghini". Ogni giorno ne guida una, e il giorno successivo ne guida una diversa, scelta a caso (eventualmente, della stessa marca). Descrivere mediante una catena di Markov lo "stato" «marca dell'auto guidata da Lewis» (1=Mercedes, 2=Ferrari, 3=Lamborghini).

a) Scrivere la matrice di transizione P della catena di Markov che descrive la marca dell'auto guidata da Lewis in un determinato giorno; dimostrare che P è *regolare*, calcolare la distribuzione *invariante* e mostrare che essa è anche *reversibile.*

b) Qual è la posta equa di una scommessa in seguito alla quale riceveremo 120€ se tra un anno da oggi Lewis starà guidando una Ferrari, e nulla altrimenti? Spiegare.

Soluzione

```
a) La matrice di transizione è p=
```

$$\begin{pmatrix} \frac{2}{5} & \frac{2}{5} & \frac{1}{5} \\ \frac{3}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{3}{5} & \frac{2}{5} & 0 \end{pmatrix}$$

Si vede facilmente che ogni stato comunica con ogni altro, quindi tutti gli stati sono ricorrenti e la matrice è irriducibile; inoltre ci sono termini non nulli sulla diagonale principale, quindi P è regolare.

Siccome P è regolare, c'è una sola distribuzione invariante, calcolata qui sotto:

```
Solve[{{π1, π2, π3}.p == {π1, π2, π3}, π1 + π2 + π3 == 1}, {π1, π2, π3}]
```

$$\left\{\left\{\pi 1 \to \frac{1}{2},\ \pi 2 \to \frac{1}{3},\ \pi 3 \to \frac{1}{6}\right\}\right\}$$

$\{\pi_1,\pi_2,\pi_3\}$ è anche *reversibile* perché soddisfa le relazioni $\pi_i \cdot p_{i,j} = \pi_j \cdot p_{j,i}$ per $i,\ j \in \{1, 2, 3\}$; la prova consiste in un calcolo diretto.

b) Tra un anno saranno trascorsi 365 giorni, quindi 365 transizioni; poiché P è regolare, qualunque sia la distribuzione iniziale v è $v \cdot P^n \longrightarrow \{\pi_1, \pi_2, \pi_3\}$ per $n \longrightarrow +\infty$; quindi $v \cdot P^{365} \approx \{\pi_1, \pi_2, \pi_3\} = \left\{\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right\}$; la probabilità che Lewis guidi una Ferrari (stato 2) è (circa) $\frac{1}{3}$; la tariffa equa per la scommessa è $\frac{1}{3} \cdot 120\,€ = 40\,€$.

08 luglio 2016 es.3 (La successione a Vittoria)

Vittoria (70 anni) è la regina di un piccolo stato, e ha due figlie, Elisabetta (45) e Maddalena (40), prima e seconda in ordine per la successione al trono. Se le due eredi moriranno prima della madre, la monarchia finirà e si passerà alla repubblica; altrimenti la corona passerà dalla madre a una delle figlie, a Elisabetta se entrambe saranno in vita alla morte della madre; Maddalena è inoltre designata quale preima erede al trono di Elisabetta.

a) Calcolare, in funzione dei valori l_x, d_x (per la popolazione femminile) la probabilità che fra tre anni la regina sia Maddalena.

b) Calcolare la probabilità che Vittoria sia l'ultima regina, ossia che sopravviva a entrambe le figlie (per semplicità escludiamo il caso in cui la morte di Vittoria avvenga nello stesso anno di almeno una delle sue figlie).

Soluzione

a) L'evento descritto si verifica se entro tre anni, non importa in quale ordine, sia Vittoria sia Elisabetta passano a miglior vita, mentre Maddalena è ancora vivente fra tre anni. I tre eventi (decesso di Vittoria e Elisabetta entro tre anni; sopravvivenza di Maddalena per almeno tre anni) sono da considerare indipendenti; la probabilità che tutti e tre si verifichino è quindi uguale a

$$\frac{l_{70}-l_{73}}{l_{70}} * \frac{l_{45}-l_{48}}{l_{45}} * \frac{l_{43}}{l_{40}}.$$

b) Tenendo presente la clausola posta "per semplicità", l'evento in oggetto si verifica se la morte di Vittoria avviene tra x anni, mentre ciascuna delle figlie vive *meno* di x anni da oggi.

La probabilità che Vittoria muoia tra x anni, cioè in età $70 + x$ è $\frac{d_{70+x}}{l_{70}}$

La probabilità che Elisabetta muoia entro $x - 1$ anni, è $\frac{l_{45}-l_{45+x}}{l_{45}}$

La probabilità che Maddalena muoia entro $x - 1$ anni, è $\frac{l_{40}-l_{40+x}}{l_{40}}$

La probabilità che tra x anni la monarchia abbia termine è il prodotto di queste tre espressioni; la probabilità che Vittoria sia l'ultima regina è la somma di questi prodotti, relativamente agli anni x nei quali non è nulla la probabilità di esistenza in vita di Vittoria:

$$\sum_{x=1}^{\omega-70} \frac{d_{70+x}}{l_{70}} * \frac{l_{45}-l_{45+x}}{l_{45}} * \frac{l_{40}-l_{40+x}}{l_{40}}.$$

08 luglio 2016 es.4 (pomodori pelati)

L'etichetta sul barattolo di pomodori pelati di una certa marca dichiara *"peso netto 400g".* Siccome si sospetta che il contenuto dei barattoli sia mediamente un po' scarso rispetto al dichiarato, apriamo quattro barattoli di pomodori e pesiamo il contenuto; i risultati sono

```
lista = {410, 370, 400, 380};
```

a) Stabilire se questi valori sono tali da respingere al livello 5%, con test unilaterale, l'ipotesi: *"Il contenuto medio dei barattoli di pomodoro è $\mu \geq 400$"*

b) La conclusione sarebbe stata la stessa nel caso che gli stessi valori per la media campionaria e la varianza corretta si fossero ottenuti da un campione di 20 barattoli anziché 4?

Soluzione

a) La media campionaria e la varianza corretta del campione valgono rispettivamente

```
x̄ = Mean[lista]; s2 = Variance[lista]; s = N[√s2]; Print[{x̄, s2, s}]
```

$\left\{390, \frac{1000}{3}, 18.2574\right\}$

La statistica-test è $T = \frac{\overline{x}-\mu}{s}\sqrt{n}$, con $n = 4$ e $\mu = 400$. T ha distribuzione di Student con 3 gradi di libertà. L'insieme di rifiuto al livello α dell'ipotesi, con test unilaterale, è $D =]-\infty, -t_{1-\alpha}(3)]$; il quantile $t_{0.95}(3)$ vale

```
Quantile[StudentTDistribution[3], 0.95]
```

2.35336

Il valore di T con i nostri dati sperimentali è

$$\frac{\overline{x} - 400}{s}\sqrt{4}$$

-1.09545

che non appartiene all'insieme di rigetto; l'ipotesi non viene respinta.

b) Se da un campione si 20 barattoli abbiamo gli stessi valori di media campionaria e varianza corretta, la statistica-test T si calcola come sopra, tranne che il fattore $\sqrt{4}$ viene sostituito con $\sqrt{20}$: si ottiene il risultato

$$\frac{\overline{x} - 400}{s}\sqrt{20}$$

-2.44949

Questa volta T ha distribuzione di Student con 19 gradi di libertà. L'insieme di rifiuto al livello α dell'ipotesi, con test unilaterale, è $D =]-\infty, -t_{1-\alpha}(19)]$; il quantile $t_{0.95}(19)$ vale

```
Quantile[StudentTDistribution[19], 0.95]
```

1.72913

Il valore sperimentale di T appartiene all'insieme di rigetto. L'ipotesi viene respinta.

08 luglio 2016 es.5 (il giocoliere al semaforo)

Carl Friedrich, giovane matematico disoccupato, intrattiene gli automobilisti che sostano al rosso di un semaforo in un incrocio molto trafficato con un piccolo spettacolo di giocoleria con clavette e altri attrezzi, e accetta gli spiccioli che alcuni automobilisti gli offrono. Nei 51 giorni scorsi l'incasso medio giornaliero è stato di 25€, con una varianza corretta di $(7\,€)^2$.

a) Calcolare, al livello 97.5%, un intervallo di confidenza superiormente illimitato per la media μ dell'incasso giornaliero e un intervallo di confidenza per la varianza, della forma $[0, h]$.

b) Carl ha deciso di mettere da parte l'incasso dei prossimi 100 giorni per finanziare le sue vacanze. Calcolare qual è l'importo minimo sul quale Carl potrà fare affidamento, con probabilità 90%.

Soluzione

a) Il testo fornisce $\overline{x} = 25$, $s = 7$, $n = 51$. L'intervallo di confidenza per μ, nei termini richiesti, è $[\overline{x} - \frac{s}{\sqrt{n}} t_{0.975}(50), +\infty[$; il quantile della distribuzione t vale

```
t = Quantile[StudentTDistribution[50], 0.975]
```

```
2.00856
```

quindi l'estremo di sinistra dell'intervallo di confidenza cercato per la media è

```
m = 25 - 7/Sqrt[51] t
```

```
23.0312
```

L'intervallo di confidenza per la varianza è $[0, h]$ con $h = \frac{(n-1)\,s^2}{\chi^2_{0.025}(n-1)}$. Il quantile $\chi^2_{0.025}(50)$ vale

```
In[1061]:= χ = Quantile[ChiSquareDistribution[50], 0.025]
```

```
Out[1061]= 32.3574
```

quindi l'estremo h dell'intervallo è

```
In[1062]:= h = (50 * 49)/χ
```

```
Out[1062]= 75.7169
```

b) Lo scarto quadratico medio σ della variabile "incasso giornaliero" è, con probabilità 0.975, non superiore a

```
In[1063]:= Sqrt[h]
```

```
Out[1063]= 8.70155
```

b) Sia X la variabile aleatoria “incasso giornaliero”, la cui media μ e varianza σ^2 non sono note. Per il Teorema Limite Centrale, la variabile Y “incasso complessivo nei prossimi 100 giorni”, somma di 100 esemplari di X indipendenti ed equidistribuiti ha approssimativamente distribuzione $N\left(100\,\mu,\ 100\,\sigma^2\right)$, e la variabile $Z = \frac{Y - 100\,\mu}{10\,\sigma}$ ha approssimativamente distribuzione $N(0, 1)$. Con probabilità 0.95 sarà $Z \geq -\phi_{0.95}$, il cui valore è

```
In[1065]:= ϕ0.95 = Quantile[NormalDistribution[], .95]; Print[-ϕ0.95]
```

```
-1.64485
```

$$Z \geq -\phi_{0.95} \text{ equivale a } Y \geq 100\,\mu - 10\,\sigma * \phi_{0.95}.$$

I valori di μ e σ sono sconosciuti; tuttavia i dati raccolti ci hanno permesso di stabilire che vi è probabilità 0.975 che sia $\mu \geq 23.03$, e ancora probabilità 0.975 che sia $\sigma \leq 8.70$ (arrotondiamo i valori alla seconda cifra decimale); con probabilità $0.975^2 \approx 0.95$ sono vere *entrambe* queste disuguaglianze. Allora, con probabilità $0.95^2 \approx 0.90$ si avrà

$$Y \geq 100 * 23.03 - 10 * 8.70 * \phi_{0.95}.$$

Il secondo membro vale

In[1066]:= $100 * 23.03 - 10 * 8.70 * \phi_{0.95}$

Out[1066]= 2159.9

Questo è l'importo che con probabilità 0.90 Carl incasserà, come minimo, nei prossimi 100 giorni di attività.

08 luglio 2016 es.6 (Le scelte degli studenti dopo il dipoma)

In una determinata città viene svolta una ricerca su 300 studenti dell'ultimo anno di scuola superiore riguardo alle scelte di studio o lavoro successive al diploma. Si vuole analizzare l'eventuale correlazione tra il tipo di scuola frequentata e le scelte successive. I dati raccolti sono i seguenti

□	Laurea Scient	Laurea Uman	Lavoro
Lic.Classico	44	25	21
Lic.Scientifico	60	30	30
Altra Scuola	26	35	29

a) Stabilire se, al livello del 5%, si può ritenere che la scelta dei giovani al termine della scuola secondaria sia indipendente dal tipo di scuola frequentata.
b) Calcolare il livello di significaticità del test mediante interpolazione lineare dei dati della tavola appropriata.

Soluzione

La seguente matrice aggiunge ai dati del problema le distribuzioni marginali relative a "Scuola" e "Dopo Scuola":

```
Print[MatrixForm[mm3]]
```

□	Laurea Scient	Laurea Uman	Lavoro	Marginale Scuola
Lic.Classico	44	25	21	90
Lic.Scientifico	60	30	30	120
Altra Scuola	26	35	29	90
DopoScuola Marginale	130	90	80	300

Questa corrisponde alle seguenti frequenze relative (accoppiate e marginali), ottenute dividendo ciascun numero per 300:

$$\begin{pmatrix} 0.147 & 0.0833 & 0.0700 & 0.300 \\ 0.200 & 0.100 & 0.100 & 0.400 \\ 0.0867 & 0.117 & 0.0967 & 0.300 \\ 0.433 & 0.300 & 0.267 & 1.00 \end{pmatrix}$$

I termini di questa matrice , siano pi[h, k] sono le stime delle probabilità congiunte (h, $k \in \{1, 2, 3\}$).

Invece la matrice con le frequenze relative accoppiate teoriche in caso di indipendenza, cioè i prodotti delle distribuzioni marginali è

$$\begin{pmatrix} 0.130 & 0.0900 & 0.0800 & 0.300 \\ 0.173 & 0.120 & 0.107 & 0.400 \\ 0.130 & 0.0900 & 0.0800 & 0.300 \\ 0.433 & 0.300 & 0.267 & 1.00 \end{pmatrix}$$

I termini di questa matrice sono $p[h] * q[k]$ (h, $k \in \{1, 2, 3\}$).

La statistica-test per l'indipendenza delle due variabili "scuola frequentata" e "scelta dopo la scuola" è

$$t = 300 \sum_{h=1}^{3} \sum_{k=1}^{3} \frac{(p[h] * q[k] - pi[h, k])^2}{p[h] * q[k]}$$

che vale, con i nostri dati

```
t = N[n Sum[Sum[(p[h] * q[k] - pi[h, k])^2 / (p[h] * q[k]), {k, 1, 3}], {h, 1, 3}]]
```

$$t = N\left[n \sum_{h=1}^{3} \sum_{k=1}^{3} \frac{(p[h] * q[k] - pi[h, k])^2}{p[h] * q[k]}\right]$$

```
11.2653
```

Se le due variabili considerate sono indipendenti, *t* ha aprossimativamente distribuzione $\chi^2(4)$ (i gradi di libertà sono (3-1)×(3-1)). Il quantile che limita inferiormente la regione di rifiuto dell'ipotesi di indipendenza è

```
qq = Quantile[ChiSquareDistribution[4], .95]
```

```
9.48773
```

L'ipotesi viene rifiutata.

b) Il livello di significatività del test che ha dato questi risultati sperimentali è α tale che $t = \chi^2_{1-\alpha}(4)$. Cerchiamo $\beta = 1 - \alpha$ in modo che siano allineati i tre punti (vedi tavola dei quantili di $\chi^2_{1-\alpha}(4)$)

(0.975, 11.1433); (0.99, 13.2767); (β, 11.2653)

Quindi deve essere

$$\frac{\beta - 0.975}{0.99 - 0.975} = \frac{11.2653 - 11.1433}{13.2767 - 11.1433}$$

da cui si ricava

$$\text{Solve}\left[\frac{\beta - 0.975}{0.99 - 0.975} == \frac{11.2653 - 11.1433}{13.2767 - 11.1433}, \beta\right]$$

$\{\{\beta \to 0.975858\}\}$

quindi $\alpha = 0.024142$.

Il valore “esatto” di α, calcolato da *Mathematica*, è

```
1 - CDF[ChiSquareDistribution[4], 11.2653]
```

```
0.0237388
```

Nota. Usando l’altra versione della statistica test, come nel libro di P.Baldi, si trova il valore

$$tt = N\left[n \sum_{h=1}^{3} \sum_{k=1}^{3} \frac{(p[h] * q[k] - pi[h, k])^2}{pi[h, k]}\right]$$

```
12.7474
```

che ancora conduce al rifiuto dell’ipotesi; cambia però il livello di significatività perché si trova ora β dalla equazione

$$\text{Solve}\left[\frac{\beta - 0.975}{0.99 - 0.975} == \frac{12.7474 - 11.1433}{13.2767 - 11.1433}, \beta\right]$$

$\{\{\beta \to 0.986278\}\}$

quindi $\alpha = 0.013722$; la statistica non è una scienza “del tutto” esatta...

In questo caso il valore “esatto” di α, calcolato da *Mathematica*, è

```
1 - CDF[ChiSquareDistribution[4], 12.7474]
```

```
0.0125783
```