01 giugno 2018 es.1 (Paperino contro Gastone)

Paperino e Gastone si sfidano nel seguente gioco: in ciascuna estrazione del Lotto (5 numeri distinti estratti tra 1 e 90) Paperino scommette sull'uscita del 13, Gastone sull'uscita di almeno due tra 1, 3, 13. La sfida sarà vinta da colui che per primo vedrà realizzata la sua scommessa; se le scommesse dei due contendenti si realizzeranno per la prima volta nella stessa estrazione, la sfida sarà conclusa in pareggio.

a) Calcolare le probabilità p, q, r che *nella prima estrazione* la sfida abbia termine, rispettivamente con la vittoria di Paperino, o di Gastone, o in pareggio.

b) Verificato che le probabilità richieste in (a) (arrotondate a due cifre significative) sono rispettivamente 0.051; 0.0024; 0.0049, calcolare la probabilità che sia Paperino ad aggiudicarsi la sfida, non necessariamente nella prima estrazione.

Soluzione

a) La vittoria di Paperino *nella prima estrazione* si realizza in caso di uscita del 13, e di nessuno dei due numeri 1, 3, perché l'uscita di uno di questi due numeri, oltre al 13, darebbe luogo al pareggio. La probabilità p di ciò si può ottenere come prodotto di tre fattori:

$\frac{5}{90}$ (probabilità che tra i 5 estratti ci sia il 13);

$\frac{85}{89}$ (probabilità che tra gli 85 non estratti ci sia 1, condizionata all'informazione che 13 fa parte degli estratti);

$\frac{84}{88}$ (probabilità che tra gli 85 non estratti ci sia 3, condizionata all'informazione che 13 fa parte degli estratti e 1 fa parte dei non estratti; rimangono quindi 88 numeri di cui non si conosce la sorte, 84 dei quali non estratti). Allora:

$$p = \frac{5}{90} * \frac{85}{89} * \frac{84}{88}; \ p_1 = \text{NumberForm}[N[p], 2];$$

```
Print[{p, N[p], p1}]
```

$$\left\{\frac{595}{11748}, 0.0506469, 0.051\right\}$$

Il calcolo di p si può svolgere anche con ragionamento esclusivamente combinatorio: tra i $\binom{90}{5}$ possibili esiti dell'estrazione, quelli favorevoli alla vittoria di Paperino sono $\binom{87}{4}$ cioè le combinazioni dei numeri diversi da 1, 3, 13 a quattro alla volta, che sono le quaterne di numeri diversi da accompagnare al 13 affinché si realizzi un'estrazione che stabilisce la vittoria di Paperino:

$$p = \frac{\text{Binomial}[87, 4]}{\text{Binomial}[90, 5]}$$

$$\frac{595}{11748}$$

La vittoria di Gastone *nella prima estrazione* si realizza in caso di uscita dei numeri 1 e 3 ma non del 13. La probabilità q di ciò si può ottenere come prodotto di tre fattori, similmente a quanto fatto sopra:

$$q = \frac{5}{90} * \frac{4}{89} * \frac{85}{88}; \ q_1 = \text{NumberForm}[N[q], 2];$$

```
Print[{q, N[q], q1}]
```

$$\left\{\frac{85}{35244}, 0.00241176, 0.0024\right\}$$

Anche per q, ragionamento combinatorio: ai numeri 1 e 3 vanno accompagnati 3 numeri qualunque tra gli 87 diversi da 1, 3, 13:

$$q = \frac{\text{Binomial}[87, 3]}{\text{Binomial}[90, 5]}$$

$$\frac{85}{35244}$$

La sfida ha termine *alla prima estrazione* con risultato di *pareggio* se esce il 13 e *almeno uno* tra 1 e 3.

La probabilità che esca il 13 è $\frac{1}{18}$; sotto questa condizione, la probabilità che esca almeno uno tra 1 e 3 è

$$\mathbf{1 - \frac{85}{89} * \frac{84}{88}}$$

$$\frac{173}{1958}$$

e quindi la probabilità *r* del pareggio alla prima estrazione è

$$\mathbf{r = \frac{1}{18} * \left(1 - \frac{85}{89} * \frac{84}{88}\right);\ r_1 = NumberForm[N[r],\ 2];}$$

```
Print[{r, N[r], r1}]
```

$$\left\{\frac{173}{35\,244},\ 0.00490864,\ 0.0049\right\}$$

Il calcolo di *r* con ragionamento combinatorio dà

$$\mathbf{r = 2 * \frac{Binomial[87,\ 3]}{Binomial[90,\ 5]} + \frac{Binomial[87,\ 2]}{Binomial[90,\ 5]}}$$

$$\frac{173}{35\,244}$$

(il primo addendo è la probabilità che siano estratti il 13 e *uno solo* tra 1 e 3; il secondo è la probabilità che 13, 1 e 3 siano estratti tutti e tre).

È utile (per la parte (b)) calcolare anche la probabilità che la prima estrazione non sia conclusiva della sfida tra Paperino e Gastone; questa è evidentemente

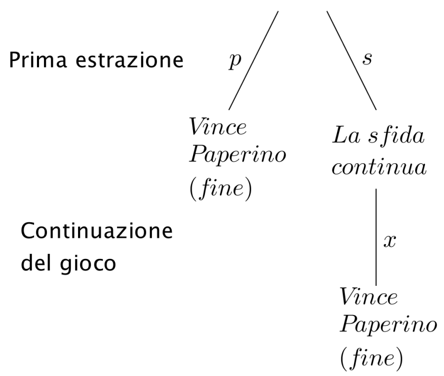
```
s = 1 - (p + q + r); s1 = NumberForm[N[s], 3];
Print[{s, N[s], s1}]
```

$$\left\{\frac{3689}{3916},\ 0.942033,\ 0.942\right\}$$

b) Detta *x* la probabilità che il vincitore della sfida sia Paperino, questa è la somma delle probabilità dei due eventi disgiunti:

"Paperino si aggiudica la sfida alla prima estrazione"

"Paperino si aggiudica la sfida in una estrazione successiva alla prima"



Il grafo ad albero riporta soltanto gli esiti favorevoli alla vittoria finale di Paperino, con le rispettive probabilità; si ha che $x = p + s * x$; da questa relazione si ricava *x*:

```
NSolve[x == p + s * x, x]
```

```
{{x → 0.873715}}
```

Se si usano gli arrotondamenti p_1, s_1 per *p* e *s* si ottiene un risutato leggermente diverso:

$$\mathbf{N\left[\frac{0.051}{1 - 0.942}\right]}$$

```
0.87931
```

Analogamente si calcolano le probabilità y, z (non richieste dal testo del problema) che la sfida sia vinta da Gastone e (rispettivamente) che finisca in parità:

```
NSolve[y == q + s * y, y]
```

```
{{y → 0.0416055}}
```

```
NSolve[z == r + s * z, z]
```

```
{{z → 0.0846794}}
```

01 giugno 2018 es.2 (Abilitazione alla professione notarile)

In un determinato Paese i laureati in Giurisprudenza che vogliono intraprendere la professione di Notaio debbono superare un esame di Stato che consiste in scritto e orale.
L'esito della prova scritta "gravemente insufficiente" comporta l'esclusione definitiva dall'ordine professionale; l'esito "insufficiente" non dà diritto ad accedere all'orale, ma consente la ripetizione dello scritto; l'esito "sufficiente" dà l'ammissione all'orale.
L'orale "gravemente insufficiente" comporta ancora l'esclusione definitiva dall'ordine professionale; l'orale "insufficiente" impone al candidato di ripetere la prova scritta ed eventualmente l'orale; l'orale "sufficiente" abilita alla professione.

Claudio, candidato all'esame, ritiene per sé ugualmente probabile che lo scritto abbia ciascuno dei tre esiti specificati, e ritiene che in un eventuale orale, la probabilità di esito "gravemente insufficiente" sia quattro volte inferiore alla probabilità di avere un esito diverso, e che le probabilità di esito "insufficiente" e "sufficiente" siano uguali

a) Descrivere il problema attraverso una catena di Markov con quattro stati: 1=escluso dall'Ordine; 2=candidato alla prova scritta; 3=ammesso all'orale; 4=abilitato. Scrivere la matrice di transizione, classificare gli stati (transitori, ricorrenti, assorbenti); dire se la catena è irriducibile e se è regolare.

b) Calcolare la probabilità che Claudio riesca a ottenere l'abilitazione alla professione notarile, dopo che ha superato la prova scritta (ossia quando si trova nello stato 3).

Soluzione

a) La compilazione della prima, seconda e quarta riga della matrice di transizione è immediata; calcoliamo i termini della terza riga.
Dallo stato 3 non si può rimanere in tale stato, quindi $p_{3,3} = 0$; si sa poi che $p_{3,0} = \frac{1}{4}(p_{3,2} + p_{3,4})$ e che $p_{3,2} = p_{3,4}$.
Inoltre deve essere $p_{3,0} + p_{3,2} + p_{3,4} = 1$. Si ricava:

$$p_{3,0} = \frac{1}{5}; \quad p_{3,2} = p_{3,4} = \frac{2}{5}$$

La matrice di transizione è pertanto:

$$\mathbf{p} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{5} & \frac{2}{5} & 0 & \frac{2}{5} \\ 0 & 0 & 0 & 1 \end{pmatrix};$$

Gli stati 1 e 4 sono assorbenti (quindi ricorrenti); gli stati 2 e 3 sono transitori, in quanto ciascuno di essi comunica (per esempio) con 4, il quale non comunica con 2 né con 3. La catena non è regolare né irriducibile.

b) indicata con C la classe chiusa $\{4\}$, ciò che è richiesto è il valore di

$$\lambda_3 = P(\exists\, n \in \mathbb{N},\ X_n = 4 \ |\ X_0 = 3)$$

Chiamato anche

$$\lambda_2 = P(\exists\, n \in \mathbb{N},\ X_n = 4 \ |\ X_0 = 2)$$

e chiamato $D = \{2, 3\}$ l'insieme degli stati transitori, si sa che (λ_2, λ_3) è soluzione del sistema lineare

$$\lambda_i = p_{i,4} + \sum_{x_j \in D} p_{i,j}\, \lambda_j\,, \quad i = 2,\ 3$$

cioè

$$\lambda_2 = \frac{1}{3} \cdot \lambda_2 + \frac{1}{3} \cdot \lambda_3 \quad ; \qquad \lambda_3 = \frac{2}{5} + \frac{2}{5} \cdot \lambda_2$$

il quale dà: $\lambda_2 = \frac{1}{4}$, $\lambda_3 = \frac{1}{2}$.

01 giugno 2018 es.3 (La pensione per Mario e Tosca)

Mario, di età 30, è un pittore, sposato con Tosca (25). Avendo venduto per 50 000 € un suo quadro, vuole investire il ricavato per assicurarsi una rendita vitalizia di importo costante, che gli sarà pagata annualmente a partire dal suo 65° compleanno. Il "premio puro" sul quale la compagnia calcolerà l'importo della rendità è la metà di quanto versato, cioè 25 000 €, trattenendo l'altra metà per tasse, spese, e per proprio guadagno.

a) Esprimere in funzione dei dati demografici e del tasso di mercato i, l'importo annuo r che Mario percepirà.

b) Mario cambia idea rispetto alla scelta di (a): desidera garantire anche a Tosca una rendita vitalizia. Precisamente, vuole per sè una rendita vitalizia come in (a); inoltre, se Tosca gli sopravvivrà, vuole che allo scadere di ogni anno che Tosca compirà da vedova, ella riceva un importo uguale alla metà di quanto percepiva lui. Vale a dire: dal 65° compleanno di Mario, e finché Mario sarà in vita, sarà pagato a lui un importo R e niente altro; se Mario morirà prima di Tosca, ogni anno Tosca riceverà $\frac{R}{2}$, finché sarà in vita.
Dare un'espressione di R, in funzione dei dati demografici e del tasso di mercato i.

Soluzione

a) La prestazione grarantita dalla Compagnia è, alla scadenza di ciascun anno a partire da 35 anni da oggi, una variabile aleatoria che vale r se Mario è in vita, 0 altrimenti. Indicato con $v = \frac{1}{1+i}$ il coefficiente di sconto annuo, la speranza matematica del valore attuale del complesso dei pagamenti è

$$V = \sum_{k=0}^{\omega-65} r \cdot \frac{l_{65+k}}{l_{30}} \cdot v^{k+35}$$

Siccome si desidera $V = 25\,000$ €, l'importo annuo che sarà pagato a Mario ammonta a

$$r = \frac{25\,000}{\sum_{k=0}^{\omega-65} \frac{l_{65+k}}{l_{30}} \cdot v^{k+35}}$$

b) In questo caso la prestazione assicurata alla scadenza di ciascun anno a partire da 35 anni da oggi è una variabile aleatoria che vale R se Mario è in vita, vale $\frac{R}{2}$ se Tosca è in vita e Mario no, vale 0 se entrambi sono morti.
La probabilità che in uno degli anni osservati Mario sia in vita è $\frac{l_{65+k}}{l_{30}}$; la probabilità che sia in vita Tosca ma non Mario è $\left(1 - \frac{l_{65+k}}{l_{30}}\right) \cdot \frac{l'_{60+k}}{l'_{25}}$.

La speranza matematica del valore attuale del complesso dei pagamenti dovuti è perciò

$$V = \sum_{k=0}^{\omega-60} \left(R \cdot \frac{l_{65+k}}{l_{30}} + \frac{R}{2} \cdot \left(1 - \frac{l_{65+k}}{l_{30}}\right) \cdot \frac{l'_{60+k}}{l'_{25}}\right) \cdot v^{k+35}$$

convenendo di porre $l_x = 0$ per $x \geq \omega$.

Siccome $V = 25\,000$ €, l'importo annuo R che sarà pagato a Mario ammonta a

$$R = \frac{25\,000}{\sum_{k=0}^{\omega-60} \left(\frac{l_{65+k}}{l_{30}} + \frac{1}{2} \cdot \left(1 - \frac{l_{65+k}}{l_{30}}\right) \cdot \frac{l'_{60+k}}{l'_{25}}\right) \cdot v^{k+35}}$$

mentre Tosca riceverà la metà di questo importo ad ogni suo compleanno, a partire dal 60°, che ella trascorrerà da vedova.

01 giugno 2018 es.4 (La statura delle ragazze sportive)

In un "Liceo sportivo" di Bologna viene misurata la statura di cinque studentesse, con i seguenti risultati (in centimetri)

```
lista = {158, 170, 179, 166, 162};
```

a) Calcolare un intervallo di confidenza al livello 95% per la "statura media" delle studentesse di quella scuola, centrato nella media campionaria.

b) Scopriamo che le misurazioni di (a) sono affette da un errore sistematico, risultando tutte errate per difetto del 4% rispetto alle misure effettive. Come cambia l'intervallo di confidenza dopo la rettifica dei dati? Fornire il risultato e spiegarne la ragione, senza ripetere tutti i calcoli come in (a).

Soluzione

a) I valori osservati della media campionaria (espressa in forma di numero decimale) e della varianza campionaria corretta sono

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

```
{167, 65, √65}
```

L'intervallo di confidenza per la media al livello del 95% (ossia $1-\alpha$ con $\alpha = 0.05$), centrato in $\overline{x}$ è

$$\left[\overline{x} - \frac{s}{\sqrt{5}}\, t_{1-\frac{\alpha}{2}},\ \overline{x} + \frac{s}{\sqrt{5}}\, t_{1-\frac{\alpha}{2}}\right]$$

con $t_{1-\frac{\alpha}{2}}$ quantile di livello $1-\frac{\alpha}{2} = 0.975$ per la distribuzione di Student con 4 gradi di libertà:

```
t_{1-α/2} = Quantile[StudentTDistribution[4], 0.975]
```

```
2.77645
```

Gli estremi dell'intervallo di confidenza cercato sono

```
Flatten[{x̄ - s/√5 t_{1-α/2}, x̄ + s/√5 t_{1-α/2}}]
```

```
{156.989, 177.011}
```

b) Le misure corrette si ottengono dividendo per 0.96 tutti i dati della `lista` (e non moltiplicandoli per 1.04!); i dati corretti sono

```
lista2 = 1/0.96 lista
```

```
{164.583, 177.083, 186.458, 172.917, 168.75}
```

La media campionaria è uguale a quella di `lista` moltiplicata per lo stesso fattore, ossia a $\frac{167}{0.96}$:

```
Print[{Mean[lista2], Mean[lista] * 1/0.96}]
```

```
{173.958, 173.958}
```

```
ȳ = Mean[lista2]
```

```
173.958
```

La varianza di `lista2`, e così pure la sua stima corretta, si ottengono da quelle di `lista` moltiplicando per $\frac{1}{(0.96)^2}$; quindi lo stimatore della varianza è, relativamente ai nuovi dati,

```
ss2 = Variance[lista]/0.96^2
```

```
70.5295
```

Controlliamo:

```
Variance[lista2]
```

```
70.5295
```

Lo stimatore dello scarto quadratico medio è

```
ss = √ss2
```
8.39819

L'intervallo di confidenza al livello del 95% con i dati così corretti si ottiene ora come in (a), sostituendo $\overline{x}$ con $\overline{y}$ e s con ss:

$$\text{Flatten}\left[\left\{\overline{y} - \frac{ss}{\sqrt{5}} t_{1-\frac{\alpha}{2}},\ \overline{y} + \frac{ss}{\sqrt{5}} t_{1-\frac{\alpha}{2}}\right\}\right]$$

{163.531, 184.386}

01 giugno 2018 es.5 (Ancora sulla statura delle studentesse)

Oltre alle misurazioni sulle allieve del Liceo Sportivo (es.3), viene misurata la statura di 8 studentesse di un Liceo Classico di Bologna; questa volta si ottiene come media 163 cm e come stima corretta della varianza 49 cm^2.

a) Stabilire con test bilaterale al livello 5% se si può ritenere che non ci sia differenza tra le statura media delle studentesse delle delle due scuole. Si faccia uso per il Liceo Sportivo dei dati originali di (a) dell'esercizio 3.

b) Ripetere il test supponendo questa volta che per le studentesse del Liceo classico si siano ottenute la media e la viarianza stimata indicate, non dalla misura di 8 studentesse, ma di 47.

Soluzione

Il test da eseguire è quello per il confronto fra le medie di campioni gaussiani indipendenti; le dimensioni dei due campioni sono $n = 5$, $m = 8$. Bisogna innanzitutto calcolare la "varianza totale". Le stime della varianza nei due campioni sono

```
sa2 = Variance[lista]; sb2 = 49;
```

e allora

$$st2 = \frac{1}{5 + 8 - 2.} ((5 - 1)\, sa2 + (8 - 1)\, sb2)$$

54.8182

```
st = √st2
```
7.40393

```
x̄ = Mean[lista]; ȳ = 163.; Print[{x̄, ȳ}]
```
{167, 163.}

La statistica-test per questo problema è

$$t = \frac{\overline{x} - \overline{y}}{st * \sqrt{\frac{1}{5} + \frac{1}{8}}}$$

0.947668

Se l'ipotesi di uguaglianza delle due medie è vera, i valori di t hanno distribuzione
di Student con $5 + 8 - 2 = 11$ gradi di libertà; la regione di rifiuto dell'ipotesi al livello α per il test bilaterale è $D =]-\infty, -t_{1-\frac{\alpha}{2}}[\cup] t_{1-\frac{\alpha}{2}}, +\infty[$, con $t_{1-\frac{\alpha}{2}}$ quantile di livello $1 - \frac{\alpha}{2}$ per la distribuzione di Student poc'anzi detta; nel nostro caso è

$$t_{1-\frac{\alpha}{2}} = \text{Quantile}[\text{StudentTDistribution}[11], 0.975]$$

2.20099

Il valore sperimentale di t non appartiene a D. L'ipotesi non viene respinta.

b) Rispetto ad (a) cambia il valore della stima totale della varianza, che adesso è

```
stt2 = 1/(5 + 47 - 2.) ((5 - 1) sa2 + (47 - 1) sb2)
```

50.28

e anche la statistica-test, che adesso è

```
tt = (x̄ - ȳ)/(Sqrt[stt2] * Sqrt[1/5 + 1/47])
```

1.19921

La regione di rifiuto dell'ipotesi appare come in (a): $D =]-\infty, -t_{1-\frac{\alpha}{2}}[\cup] t_{1-\frac{\alpha}{2}}, +\infty[$, dove però $t_{1-\frac{\alpha}{2}}$ è ora il quantile di livello $1-\frac{\alpha}{2}$ per la distribuzione di Student con $5+47-2=50$ gradi di libertà:

```
t1-α/2 = Quantile[StudentTDistribution[50], 0.975]
```

2.00856

Anche in questo caso non ci sono elementi sufficienti per rifiutare l'ipotesi.

Commento. La taglia molto maggiore di uno dei due campioni, passando da 8 a 47 elementi, non ha cambiato di molto le cose. Questo si spiega qualitativamente osservando che per avere un cambiamento importante sarebbero dovuti divenire più numerosi *entrambi* i campioni; le espressioni di t e di tt rispettivamente in (a) e (b) mostrano infatti al denominatore $\sqrt{\frac{1}{n}+\frac{1}{m}}$, che vale la prima volta $\sqrt{\frac{1}{5}+\frac{1}{8}}$, la seconda $\sqrt{\frac{1}{5}+\frac{1}{47}}$; c'è una diminuzione del radicando, che è una delle cause per cui è $tt > t$; ma questo incremento non è molto rilevante, perché l'addendo $\frac{1}{5}$ mantiene invariato l'ordine di grandezza della somma.

01 giugno 2018 es.6 (Studenti ribelli alla prova Invalsi)

Una determinata prova Invalsi consiste di 36 quesiti a risposta chiusa, con scelta tra 4 risposte possibili per ciascuna domanda. Uno statistico che controlla i risultati di una certa Scuola ha l'impressione che in due classi, complessivamente 50 scolari, le risposte siano state date casualmente, senza neppure leggere le domande. Per confermare (o smentire) questo sospetto egli suddivide le domande in blocchi di 4 (dalla 1 alla 4, dalla 5 alla 8, ecc.) e conta in quanti dei $\frac{50*36}{4} = 450$ gruppi di 4 domande vi sono 0, 1, 2, 3, 4 risposte esatte. I valori sono qui riassunti:

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 130 & 199 & 97 & 20 & 4 \end{pmatrix}$$

cioè 130 gruppi di 4 domande contengono 0 risposte esatte, 199 ne contengono una, ecc. Stabilire con un opportuno test se si deve oppure no respingere, al livello del 5%, l'ipotesi che le risposte siano state date a caso.

Soluzione

Se l'ipotesi "risposte a caso" è vera, le frequenze relative delle risposte esatte ai 450 gruppi di 4 seguono una distribuzione binomiale $B\left(450, \frac{1}{4}\right)$, che dà i seguenti valori:

```
p[k_] := PDF[BinomialDistribution[4, 1/4], k];
Table[p[k], {k, 0, 4}]
```

$\left\{\frac{81}{256}, \frac{27}{64}, \frac{27}{128}, \frac{3}{64}, \frac{1}{256}\right\}$

```
N[%, 3]
```

{0.316, 0.422, 0.211, 0.0469, 0.00391}

I valori sperimentali delle frequenze relative sono invece

```
N[Table[q[k]/450, {k, 0, 4}], 3]
```

```
{0.289, 0.442, 0.216, 0.0444, 0.00889}
```

La statistica-test per questo problema è

$$t = 450. * \sum_{k=0}^{4} \frac{1}{p[k]} * \left(\frac{q[k]}{450} - p[k]\right)^2$$

```
4.48077
```

Se l'ipotesi è vera, *t* valorizza una variabile la cui distribuzione è bene approssimata da $\chi^2(4)$ (Chi quadrato con $5 - 1 = 4$ gradi di libertà). Una regione di rifiuto al livello del 5% è $\left]\chi^2_{0.90}(4), +\infty\right[$. Il quantile che interessa è

```
In[1]:= Quantile[ChiSquareDistribution[4], 0.95]
Out[1]= 9.48773
```

L'ipotesi non viene respinta, vale a dire che è non è infondato il sospetto che le risposte siano state date a caso.