Lab 4

4.1 Urne e palline ancora.

L'urna **A** contiene 8 palline rosse e 7 palline nere, mentre l'urna **B** ne contiene 9 rosse e 9 nere. Viene lanciato un dado: se esce 1, estraggo (con reinserimento) 7 palline dall'urna **A**, altrimenti, estraggo (con reinserimento) 7 palline dall'urna **B**.

1. Calcolare la probabilità che tra le palline estratte ce ne siano esattamente 4 nere.
2. Supponendo che vengano estratte 5 palline nere, calcolare la probabilità che l'estrazione sia stata fatta dall'urna B.

4.1 Soluzione

Creiamo gli eventi:

- Evento $\mathcal{A}$: dopo lanciare il dado otteniamo 1 (e quindi, estraiamo le palline dall'urna A).
- Evento $\mathcal{B}$: il lancio del dado dà un risultato diverso da 1. (estrazione dall'urna B).
- Evento $\mathcal{C}$: Estraggo 4 palline nere
- Evento $\mathcal{D}$: Estraggo 5 palline nere

1. **Cosa mi viene chiesto?** $P(\mathcal{C})$. Per il teorema della prob. totale:

$$P(\mathcal{C}) = P(\mathcal{C}\,|\mathcal{A})\,P(\mathcal{A}) + P(\mathcal{B}\,|\mathcal{A})\,P(\mathcal{B})$$

Inoltre, dato che non viene esplicitamente detto che il dado è truccato, posso assumere che:
$P(\mathcal{A}) = \frac{1}{6}$ e $P(B) = \frac{5}{6}$
Poi, essendo le estrazioni fatte con reinserimento, posso creare delle v.a., che indichino il numero di palline **nere** estratte a valle delle sette estrazioni da ciascuna delle due urne, rispettivamente. Queste v.a. saranno distribuite come la v.a. Binomiale:

$$\mathcal{X}_{urnaA} \sim Bin\left(n = 7, \frac{7}{15}\right)$$

$$\mathcal{X}_{urnaB} \sim Bin\left(n = 7, \frac{9}{18}\right)$$

da cui mi rendo conto che:

$$P\left(\mathcal{C}\,|\mathcal{A}\right) = P\left(\mathcal{X}_{urnaA} = 4\right) = \binom{7}{4}\left(\frac{7}{15}\right)^4\left(\frac{8}{15}\right)^3$$

$$P\left(\mathcal{C}\,|\mathcal{B}\right) = P\left(\mathcal{X}_{urnaB} = 4\right) = \binom{7}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^3$$

e quindi la risposta al primo punto è:

$$P\left(\mathcal{C}\right) = \binom{7}{4}\left(\frac{7}{15}\right)^4\left(\frac{8}{15}\right)^3\left(\frac{1}{6}\right) + \binom{7}{4}\left(\frac{1}{2}\right)^7\left(\frac{5}{6}\right)$$

(A fini dell'esame, posso lasciare la risposta espressa in questo modo... è sempre una forma esatta, e, cosa non poco importante, il risultato è supportato dal procedimento).

2. **Cosa mi viene chiesto?** $P(\mathcal{B}|\mathcal{D})$

   Dal **Teorema di Bayes**, io so che:

$$P(\mathcal{B}|\mathcal{D}) = \frac{P(\mathcal{B}|\mathcal{D}) \cdot P(\mathcal{B})}{P(\mathcal{D})}$$

   Io so anche che $P(\mathcal{B}) = \frac{5}{6}$, e che:

$$P(\mathcal{B}|\mathcal{D}) = P(\mathcal{X}_{urnaB} = 5) = \binom{7}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^2 = \binom{7}{5}\left(\frac{1}{2}\right)^7$$

   Infine, ragionando in modo simile al punto 1, avrò che:

$$P\left(\mathcal{D}\right) = \binom{7}{4}\left(\frac{7}{15}\right)^5\left(\frac{8}{15}\right)^2\left(\frac{1}{6}\right) + \binom{7}{5}\left(\frac{1}{2}\right)^7\left(\frac{5}{6}\right)$$

   Potrei dunque "montare" tutti i termini per esprimere $P(\mathcal{B}|\mathcal{D})$, ma, avendo specificato la teoria che sto usando (teorema di Bayes), avendo dato una formula per la risposta, avendo specificato il significato dei termini della formula e, infine, avendo espresso il valore esatto di ogni termine dell'equazione, l'esercizio può essere considerato concluso.

4.2 Campionamento

Sia dato un campione di ampiezza $n = 100$ estratto da una popolazione descrita da una v.a. $\mathcal{X}$ la cui deviazione standard sappiamo essere $\sigma = 4$. La media campionaria è pari a $\bar{\mathcal{X}} = 20$. E vogliamo avere una stima del vero valore atteso $\mu = \mathbb{E}[\mathcal{X}]$ con un certo grado di confidenza.

1. Quanto deve valere k per poter dire che: $P(|\bar{\mathcal{X}} - \mu| < k) = 0.95$?

2. Quanto avrebbe dovuto valere n per poter dire che: $P(|\bar{\mathcal{X}} - \mu| < \frac{1}{25}) = 0.95$?

Suggerimento: Visto che non abbiamo calcolatrice, può essere comodo sapere che:

$$\frac{1.95}{2} = 0.975$$

4.2 Soluzione

1. Per il **Teorema del limite Centrale**, io so che la media campionaria, $\bar{\mathcal{X}}$, segue una distribuzione Gaussiana (o *Normale*), con valore atteso $\mu_{\bar{\mathcal{X}}} = \mu$ (cioè, il valore atteso di $\bar{\mathcal{X}}$ è lo stesso di quello di $\mathcal{X}$), e di deviazione standard $\sigma_{\bar{\mathcal{X}}} = \frac{\sigma}{\sqrt{n}}$ (cioè, la deviazione standard di $\bar{\mathcal{X}}$ è n volte più piccola di quella di $\mathcal{X}$. In altre parole:

   $$\bar{\mathcal{X}} \sim Norm(\mu, \frac{\sigma}{\sqrt{n}})$$

   da cui avrò che, la variabile standardizzata $\frac{\bar{\mathcal{X}}-\mu}{\frac{\sigma}{\sqrt{n}}}$, seguirà una distribuzione, appunto, normale standard:

   $$\frac{\bar{\mathcal{X}} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim Norm(0, 1)$$

   Oppure, esprimendolo in modo più comodo:

   $$\frac{(\bar{\mathcal{X}} - \mu) \cdot \sqrt{n}}{\sigma} \sim Norm(0, 1)$$

   Da cui posso dire che, se la domanda richiesta si potesse esprimere in funzione di questa nuova variabile aleatoria standardizzata, potremmo tirar fuori dei valori dalla tabella $\mathbf{\Phi}$.

   Prima facciamo qualche manipolazione conveniente:

   $$P(|\bar{\mathcal{X}} - \mu| < k) = 0.95$$

   Verrà ora espressa come:

   $$P(-k < \bar{\mathcal{X}} - \mu < k) = 0.95$$

   E, poi, moltiplicando tutti i lati della diseq. per $\frac{\sqrt{n}}{\sigma}$, otteniamo:

   $$P(-k \cdot \frac{\sqrt{n}}{\sigma} < (\bar{\mathcal{X}} - \mu) \cdot \frac{\sqrt{n}}{\sigma} < k \cdot \frac{\sqrt{n}}{\sigma}) = 0.95$$

   Da cui mi rendo conto che al centro della disequazione ho una v.a. normale standard. Questa, per comodità la posso chiamare $\mathcal{Z}$, e così:

   $$P(-k \cdot \frac{\sqrt{n}}{\sigma} < \mathcal{Z} < k \cdot \frac{\sqrt{n}}{\sigma}) = \mathbf{\Phi}(k \cdot \frac{\sqrt{n}}{\sigma}) - \mathbf{\Phi}(-k \cdot \frac{\sqrt{n}}{\sigma}) = 0.95$$

Da cui, usando le regole della funzione di ripartizione, abbiamo:

$$\mathbf{\Phi}(k \cdot \frac{\sqrt{n}}{\sigma}) - [1 - \mathbf{\Phi}(k \cdot \frac{\sqrt{n}}{\sigma})] = 0.95$$

$$2\mathbf{\Phi}(k \cdot \frac{\sqrt{n}}{\sigma}) - 1 = 0.95$$

$$\mathbf{\Phi}(k \cdot \frac{\sqrt{n}}{\sigma}) = \frac{1.95}{2} = 0,975$$

A questo punto possiamo fare il lookup in modo inverso nella nostra tabella $\mathbf{\Phi}$, e capire che:

$$k \cdot \frac{\sqrt{n}}{\sigma} = 1.96$$

e cioè:

$$k \cdot \frac{\sqrt{100}}{4} = k \cdot \frac{10}{4} = k \cdot 2.5 = 1.96$$

Quindi, all'esame la soluzione al primo punto può essere lasciata espressa come:

$$k = \frac{1.96}{2.5}$$

2. Facendo lo stesso ragionamento del punto precedente, e cioè, forti del fatto che, in questo caso, **grazie al TLC**, abbiamo che:

$$P(|\bar{\mathcal{X}} - \mu| < k) = 0.95 \iff k \cdot \frac{\sqrt{n}}{\sigma} = 1.96$$

per qualsiasi k e qualsiasi n, allora possiamo risolvere l'esercizio usando:

$$\frac{1}{25} \cdot \frac{\sqrt{n}}{4} = 1.96$$

e cioè:

$$\frac{\sqrt{n}}{100} = 1.96$$

quindi:

$$n = 196^2$$

4.3 Inferenza

Il rendimento degli studenti in due corsi differenti segue una distribuzione normale. Nel corso **A**, i voti hanno un valore atteso di 75 con una deviazione standard di 5, mentre nel corso **B** il valore atteso è di 80 con una deviazione standard di 8.
Se uno studente ottiene un voto di 78, qual è la probabilità che appartenga al corso **A**?
Nota: Si assuma che, *a priori* del voto, uno studente preso a caso dall'insieme di tutti gli studenti abbia la stessa probabilità di appartenere al corso **A** che di appartenere al corso **B**.

4.3 Soluzione

1. Denotiamo gli eventi e le eventuali varibili aleatorie in gioco:

- $\mathcal{A}$: evento di appartenenza al corso **A**;
- $\mathcal{B}$: evento di appartenenza al corso **B**;
- $\mathcal{C}$: evento ”lo studente ha avuto un rendimento di 78”.

Se usiamo le variabili aleatorie $\mathcal{X}$ e $\mathcal{Y}$ per indicare il rendimento di uno studente che frequenta il corso **A** e **B**, rispettivamente, allora le informazioni formite sono:

- Nel corso A: $\mathcal{X} \sim \mathcal{N}(\mu_A = 75, \sigma_A = 5)$
- Nel corso B: $\mathcal{Y} \sim \mathcal{N}(\mu_B = 80, \sigma_B = 8)$

2. Cosa mi chiede l'esercizio?

$$\boldsymbol{P(\mathcal{A}|\mathcal{C})}$$

3. Per risolvere, utilizziamo il **Teorema di Bayes**:

$$P(\mathcal{A}|\mathcal{C}) = \frac{P(\mathcal{C}|\mathcal{A}) \cdot P(\mathcal{A})}{P(\mathcal{C})}$$

Dalla **nota** possiamo assumere che le probabilità *a priori* sugli eventi di appartenenza siano uniformi:

$$P(\mathcal{A}) = P(\mathcal{B}) = 0.5$$

Poi, il termine $P(\mathcal{C}|\mathcal{A})$, coincide proprio con:

$$P(\mathcal{C}|\mathcal{A}) = P(\mathcal{X} = 78)$$

Usando l'approssimazione della probabilità in funzione della densità di probabilità continua, sappiamo che:

$$P(\mathcal{X} = 78) = \phi_{75,5}(78) \cdot \epsilon$$

dove ϵ è un valore piccolo a piacere, e $\phi_{\mu,\sigma}(78)$ è **la densità di probabilità** di $\mathcal{X}$, cioè:

$$\phi_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

A questo punto ci manca specificare il valore di $P(\mathcal{C})$, per il quale usiamo il **teorema della probabilità totale**:

$$P(\mathcal{C}) = P(\mathcal{C}|\mathcal{A})P(\mathcal{A}) + P(\mathcal{C}|\mathcal{B})P(\mathcal{B})$$

(è **importante specificare** questo all'esame se poi si vuole usare la notazione $\phi_{\mu,\sigma}$ per esprimere il risultato finale)

Ora, facendo un ragionamento analogo a quando fatto per $P(\mathcal{C}|\mathcal{A})$, sappiamo che:

$$P(\mathcal{C}|\mathcal{B}) = P(\mathcal{Y} = 78) = \phi_{80,8}(78) \cdot \epsilon$$

Per cui, la risposta all'esercizio diviene:

$$P(\mathcal{A}|\mathcal{C}) = \frac{\phi_{75,5}(78) \cdot \epsilon \cdot 0.5}{\phi_{75,5}(78) \cdot \epsilon \cdot 0.5 + phi_{80,8}(78) \cdot \epsilon \cdot 0.5}$$

da cui osservo che posso semplificare ed esprimere la risposta senza la costante ϵ:

$$P(\mathcal{A}|\mathcal{C}) = \frac{\phi_{75,5}(78)}{\phi_{75,5}(78) + \phi_{80,8}(78)}$$

Questa è una risposta valida all'esame. Si potrebbe anche aver sostituito i valori usando la funzione di distribuzione, per semplificare ancora i termini, ma va bene così. Abbiamo dimostrato che sappiamo fare inferenza, che sappiamo usare l'approssimazione di probabilità in funzione della densità continua quando ciò è conveniente, e che sappiamo calcolare la probabilità totale.