Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 12 giugno 2017	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e di 2 ore e 30 minuti. Durante la prova e possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu passaggi di calcolo, non sara preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara eventualmente rappresentato in quest'ultima forma).
- SOLO PER STUDENTI DEL CORSO ONLINE (VECCHIO ORDINAMENTO, Prof. Gianini): Indicare quale Modulo viene svolto al fine di completare una prova d'esame parziale precedentemente sostenuta con esito positivo
  - Modulo 1 ☐
  - Moduli 2 e 3 ☐

Modulo 1

Esercizio 1.
Da un mazzo di 52 carte (ossia tredici per ognuno dei quattro semi) se ne sceglie una a caso.

(a) Quanto vale la probabilitá di estrarre una figura e un fiori?

Soluzione:
L' evento "figura" non influisce sulla probabilitá dell'evento "fiori", per cui essi sono statisticamente indipendenti. Ne segue:

$$P(\{\text{figura}\} \cap \{\text{fiori}\}) = P(\{\text{figura}\})P(\{\text{fiori}\}) = \frac{12}{52} \cdot \frac{13}{52} = \frac{3}{52}$$

(b) Quanto vale la probabilitá di estrarre una figura o una carta di fiori?

Soluzione:
Usando il risultato precedente:

$$P(\{\text{figura}\} \cup \{\text{fiori}\}) = P(\{\text{figura}\}) + P(\{\text{fiori}\}) - P(\{\text{figura}\} \cap \{\text{fiori}\}) =$$

$$\frac{12}{52} + \frac{13}{52} - \frac{3}{52} = \frac{11}{26}$$

Esercizio 2. Nel lancio ripetuto di due dadi non truccati, sappiamo che la somma dei risultati é un numero pari..

(a) Quanto vale la probabilitá di aver totalizzato 8?

Soluzione:
La probabilitá di aver totalizzato 8 é:

$$P(\{8\}) = P(\{6+2\} \cup \{5+3\} \cup \{4+4\} \cup \{3+5\} \cup \{2+6\}) = \frac{5}{36}$$

Sapendo che é uscito un numero pari, poiché $\{8\} \subset \{\text{pari}\}$ si ha:

$$P(\{8\} \mid \{\text{pari}\}) = \frac{P(\{8\} \cap \{\text{pari}\})}{P(\{\text{pari}\})} = \frac{P(\{8\})}{0.5} = \frac{5}{18}$$

Esercizio 3. Si utilizza un prodotto fornito in percentuali uguali da due ditte A e B. E' stato calcolato che, scelto a caso un esemplare difettoso, la probabilitá che esso sia stato fornito dalla ditta A vale 0.25.

(a) Se la produzione del prodotto da parte della ditta A ha un difetto di qualitá del 5%, qual é il difetto di qualitá nella produzione della ditta B?

Soluzione: (Teorema di Bayes)
I dati del problema ci dicono che

$$P(A \mid \mathsf{difettoso}) = 0.25$$

$$P(\mathsf{difettoso} \mid A) = 0.05$$

Vogliamo calcolare $P(\mathsf{difettoso} \mid B)$.
Usando Bayes:

$$P(A \mid \mathsf{difettoso}) = \frac{P(\mathsf{difettoso} \mid A)P(A)}{P(\mathsf{difettoso} \mid A)P(A) + P(\mathsf{difettoso} \mid B)P(B)} = 0.25. \tag{1}$$

Risolvendo l'equazione (1) rispetto alla probabilitá richiesta:

$$P(\mathsf{difettoso} \mid B)P(B) = \frac{0.05}{0.25} - 0.05 = 0.15$$

Il difetto di qualitá nella produzione della ditta B ṕari al 15%

Moduli 2 e 3

Esercizio 1. Una variabile aleatoria X ha densitá

$$f_X(x) = \begin{cases} x - \frac{1}{4}x^3, & per \quad 0 \leq x \leq 2 \\ 0 & \quad altrove \end{cases}$$

(a) Determinare valore atteso, varianza e mediana di X.

Soluzione:

$$E[X] = \int_0^2 x\left(x - \frac{1}{4}x^3\right) dx = \left[\frac{x^3}{3} - \frac{x^5}{20}\right]_0^2 = \frac{16}{15}$$

La varianza é $\sigma_X^2 = E[X^2] - (E[X])^2 = E[X^2] - \left(\frac{16}{15}\right)^2$, dove

$$E[X^2] = \int_0^2 x^2\left(x - \frac{1}{4}x^3\right) dx = \left[\frac{x^4}{4} - \frac{x^6}{24}\right]_0^2 = \frac{4}{3}$$

Pertanto,

$$\sigma_X^2 = \frac{4}{3} - \left(\frac{16}{15}\right)^2 = 16\left(\frac{1}{12} - \frac{16}{225}\right) \approx 0.195$$

Per calcolare la mediana x_{med}, si impone $F_X(x_{med}) = \frac{1}{2}$, dunque:

$$F_X(x_{med}) = \int_0^{x_{med}} \left(x - \frac{1}{4}x^3\right) dx = \left[\frac{x^2}{2} - \frac{x^4}{16}\right]_0^{x_{med}} = \frac{1}{2}\left(x_{med}^2 - \frac{x_{med}^4}{16}\right) = \frac{1}{2}$$

Si deve pertanto risolvere l'equazione

$$x_{med}^4 - 8x_{med}^2 + 8 = 0$$

cercando la radice che appartiene all'intervallo $0 \leq x \leq 2$.
Ponendo $t = x_{med}^2$

$$t = \begin{cases} 4 + 2\sqrt{2} \\ 4 - 2\sqrt{2} \end{cases}$$

da cui

$$\begin{cases} x_{med}^{(1,2)} = \pm\sqrt{4 + 2\sqrt{2}} \approx \pm 2.613 \\ x_{med}^{(3,4)} = \pm\sqrt{4 - 2\sqrt{2}} \approx \pm 1.0924 \end{cases}$$

Scartando le soluzioni $x_{med}^{(1,2)}$ giacché fuori dell'intervallo $0 \leq x \leq 2$, l'unica soluzione ammissibile é

$$x_{med}^{(3)} = \sqrt{4 - 2\sqrt{2}} \approx 1.0924$$

Esercizio 2.
Il giocatore A lancia un dado non truccato per 4 volte, e vince se esce almeno una volta il 6. Il giocatore B lo lancia 8 volte, e vince se il 6 esce almeno due volte.

(a) Chi ha maggiore probabilitá di vincere e perché?

Soluzione: (Distribuzione binomiale) In ogni lancio la probabilitá che esca il 6 vale $p = 1/6$ (equiprobabilitá di 6 eventi). La probabilitá di avere $X = 0$ successi in $n = 4$ prove indipendenti vale, si determina in base alla distribuzione binomiale:

$$Bin\left(X = 0 \mid n = 4, p = \frac{1}{6}\right) = \binom{4}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^4 = \left(\frac{5}{6}\right)^4 \approx 0.48226$$

Dunque, la probabilitá di vittoria per A é

$$P(A) = 1 - 0.48226 \approx 0.51774$$

Per il giocatore B, la probabilitá di avere non piú di $X = 1$ successi in $n = 8$ prove (perdendo cosí la scommessa) é

$$P_X(0 \le X \le 1) = Bin\left(X = 0 \mid n = 8, p = \frac{1}{6}\right) + Bin\left(X = 1 \mid n = 8, p = \frac{1}{6}\right)$$

ovvero

$$P_X(0 \le X \le 1) = \binom{8}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^8 + \binom{8}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^7 \approx 0.6046$$

Pertanto, la probabilitá di vittoria per B é

$$P(B) = 1 - P_X(0 \le X \le 1) \approx 0.3936$$

Esercizio 3. Si dispone di un campione di 100 misure di una variabile temporale X di una popolazione di cui non conosciamo la distribuzione, ma la cui deviazione standard é nota e vale $\sigma_X = 120$ secondi. .

(a) Qual é la probabilitá che la media campionaria differisca per piú di 3 secondi dal valore atteso teorico (incognito) dei tempi misurati?

Soluzione:
Siano:

$$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

la media campionaria;

$$\mu_X = E[X]$$

il valore atteso teorico.
Il problema ci chiede di determinare la probabilitá

$$P_X\left(\left|\frac{S_n}{n} - \mu\right| > 3\right)$$

Non conosciamo la legge di probabilitá secondo cui si distribuisce la popolazione, ma il Teorema Centrale Limite (TCL) ci assicura che per n sufficientemente grande la distribuzione degli scarti fra media campionaria $\frac{S_n}{n}$ e valore atteso teorico μ_X segue una legge Gaussiana:

$$P\left(a \le \frac{\sqrt{n}}{\sigma_X} \cdot \left(\frac{S_n}{n} - \mu\right) \le b\right) \approx \Phi(b) - \Phi(a).$$

Nel nostro caso,

$$\frac{\sigma_X}{\sqrt{n}} = \frac{120}{\sqrt{100}} = 12$$

é la deviazione standard del campione.
Usando il TCL con $a = -3, b = 3$ e passando alla variabile standardizzata $Z_n \sim N(0,1)$

$$Z_n = \frac{\frac{S_n}{n} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}},$$

il problema iniziale si riduce a calcolare la seguente:

$$P_X\left(\left|\frac{S_n}{n} - \mu_X\right| > 3\right) = P_Z\left(\frac{\sigma_X}{\sqrt{n}}|Z_n| > 3\right) = P_Z(|Z_n| > 0.25).$$

Nel caso Gaussiano, $P_Z(|Z_n| > z) = 2(1 - \Phi(z))$, dunque :

$$P_Z(|Z_n| > 0.25) = 2(1 - \Phi(0.25)).$$

Usando le tabelle della CDF standard Φ leggiamo che

$$\Phi(0.25) = 0.5987.$$

Pertanto,

$$P_X\left(\left|\frac{S_n}{n} - \mu\right| > 3\right) = 2(1 - 0.5987) = 0.8026.$$