Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 6 settembre 2016	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. (Principio di additivitá generalizzata)
Un sistema é formato da tre componenti disposti in parallelo. Il sistema funziona se almeno due dei tre componenti funzionano correttamente. Indicati con C_1, C_2, C_3 gli eventi "funzionamento corretto" del primo, secondo e terzo componente, rispettivamente, siano $P(C_1) = P(C_2) = P(C_3) = p$ con $p = 0.7$. Si assuma che gli eventi C_1, C_2, C_3 siano indipendenti tra di loro.

(a) Si mostri che la probabilitá che il sistema funzioni correttamente é maggiore della probabilitá di funzionamento dei singoli componenti.

Soluzione:
Definiamo l'evento $C =$ `"il sistema funziona correttamente"`. Il problema si risolve calcolando la probabilitá $P(C)$ e mostrando che $P(C) > p$.
Il sistema funziona se almeno due componenti funzionano correttamente. I casi possibili che si possono presentare sono 8, di questi, quelli in cui almeno due generatori sono attivi sono 4: il caso in cui funzionano tutti e tre, e i tre casi in cui si ha il malfunzionamente di uno solo. Dunque

$$C = \{C_1 \cap C_2 \cap C_3\} \cup \{\sim C_1 \cap C_2 \cap C_3\} \cup \{C_1 \cap \sim C_2 \cap C_3\} \cup \{C_1 \cap C_2 \cap \sim C_3\}.$$

Applicando il Principio di additivitá generalizzata, la probabilitá di corretto funzionamento del sistema é

$$P(C) = P(\{C_1 \cap C_2 \cap C_3\}) + P(\{\sim C_1 \cap C_2 \cap C_3\}) + P(\{C_1 \cap \sim C_2 \cap C_3\}) + P(\{C_1 \cap C_2 \cap \sim C_3\})$$

Infine, usando l'indipendenza degli eventi:

$$P(T) = p^3 + 3p^2(1-p) = 3p^2 - 2p^3 = 1.47 - 0.686 = 0.784 > p$$

Esercizio 2. (Probabilitá condizionata, regola del prodotto)
Un'urna contiene 1 pallina nera (N) e 2 palline bianche (B). Si estrae casualmente una pallina dall'urna e, dopo averne osservato il colore, la si rimette nell'urna aggiungendo altre 2 palline del colore estratto e 3 palline del colore non estratto.

i	Nere	Bianche
1	1	2
2	4	4
3	6	7
4	8	10

(a) Determinare la probabilitá che in 4 estrazioni successive, effettuate secondo la regola sopra stabilita, si ottenga la stringa (ordinata) $BNNB$

Soluzione: Indichiamo con B_i, N_i ($i = 1, \cdots, 4$) gli eventi: `"si ha una pallina Bianca (Nera) alla i-esima estrazione"`. Dopo ogni estrazione cambia lo spazio campione, e se gli esiti delle prime tre estrazioni seguono la sequenza voluta $B_1N_2N_3$, il numero delle palline presenti nell'urna quando avviene la i-esima estrazione si modifica come mostrato nella tabella.
Allora si ha:

$$P(B_1) = \frac{2}{3}$$

$$P(N_2 \mid B_1) = \frac{4}{8} = \frac{1}{2}$$

$$P(N_3 \mid N_2 \cap B_1) = \frac{6}{13}$$

$$P(B_4 \mid N_3 \cap N_2 \cap B_1) = \frac{10}{18} = \frac{5}{9}$$

e di conseguenza la probabilitá che si verifichi la sequenza $BNNB$ vale, per la regola del prodotto:

$$P(B_1 \cap N_2 \cap N_3 \cap B_4) = \frac{2}{3} \times \frac{1}{2} \times \frac{6}{13} \times \frac{5}{9} = \frac{10}{117} \approx 0.08547$$

Esercizio 3. (Teorema di Bayes)
Quattro arcieri A, B, C, D scoccano la loro freccia contemporaneamente e hanno probabilitá, rispettivamente, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$ e $\frac{1}{5}$ di colpire il bersaglio

(a) Che probabilitá c'é che dopo il tiro simultaneo risulti conficcata nel bersaglio esattamente una freccia?

Soluzione: Definiamo i seguenti eventi

- C_A = `"A colpisce il bersaglio"`
- C_B = `"B colpisce il bersaglio"`
- C_D = `"C colpisce il bersaglio"`
- C_C = `"D colpisce il bersaglio"`
- C_{una} = `"Esattamente 1 freccia nel bersaglio"`

Dai dati del problema:

$$P(C_A) = \frac{1}{2}, P(C_B) = \frac{1}{3}, P(C_C) = \frac{1}{4}, P(C_D) = \frac{1}{5}$$

Indichiamo con $\sim C_A, \cdots$ gli eventi complementari, di probabilitá $P(\sim C_A) = 1 - P(C_A), \cdots$, e per semplicitá usiamo la notazione $P(\{X \cap Y\}) \equiv P(X,Y)$. Pertanto:

$$P(C_{una}) = P(C_A, \sim C_B, \sim C_C, \sim C_D) + P(\sim C_A, C_B, \sim C_C, \sim C_D) + P(\sim C_A, \sim C_B, C_C, \sim C_D) + P(\sim C_A, \sim C_B, \sim C_C, C_D) =$$

$$= \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} \times \frac{4}{5} + \frac{1}{2} \times \frac{1}{3} \times \frac{3}{4} \times \frac{4}{5} + \frac{1}{2} \times \frac{2}{3} \times \frac{1}{4} \times \frac{4}{5} + \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} \times \frac{1}{5} = \frac{50}{120}$$

(b) Se dopo il tiro simultaneo risulta conficcata nel bersaglio una e una sola freccia, che probabilitá c'é che si tratti di quella dell'arciere A?

Soluzione: Per rispondere al quesito si tratta di inferire usando la regola di Bayes:

$$P(C_A \mid U) = \frac{P(U \mid C_A) P(C_A)}{P(U)}$$

dove abbiamo definito l'evento U = `"una e una sola freccia conficcata nel bersaglio"`.
Poiché

$$P(U \mid C_A) = P(\sim C_B, \sim C_C, \sim C_D)$$

e

$$P(U \mid C_A)P(C_A) = P(\sim C_B, \sim C_C, \sim C_D)P(C_A) = P(C_A, \sim C_B, \sim C_C, \sim C_D),$$

dove si é applicata la proprietá di indipendenza, ovvero il fatto che le prestazioni di B, C e D in un dato tiro non sono condizionate dall'esito di A (formalmente, $P(U \mid C_A) = P(\sim C_B, \sim C_C, \sim C_D \mid C_A) = P(\sim C_B, \sim C_C, \sim C_D)$).
Valendo la medesima proprietá nei casi $P(U \mid C_B), \cdots$, risulta che

$$P(U) = P(C_{una}) = \frac{50}{120}$$

calcolata precedentemente, come era da aspettarsi in termini logici (`"Esattamente 1 freccia nel bersaglio"` $\equiv$ `"una e una sola freccia conficcata nel bersaglio"`).
Sostituendo in Bayes:

$$P(C_A \mid U) = \frac{P(U \mid C_A)P(C_A)}{P(U)} = \frac{\frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} \times \frac{4}{5}}{\frac{50}{120}} = \frac{\frac{24}{120}}{\frac{50}{120}} = \frac{24}{50} = 0.48$$

.

Moduli 2 e 3

Esercizio 1. (Distribuzione Binomiale)
Viene trasmesso un segnale binario, dove ciascun bit, in maniera indipendente, nel 10% dei casi viene ricevuto erroneamente.

(a) Se ciascun pacchetto é composto di 10 bit, qual é la distribuzione del numero di errori in ricezione ?

Soluzione: Per indipendenza, la distribuzione del numero di errori X segue una legge binomiale

$$P(X) = Bin(X; n, p) = Bin(X; 10, 0.1) = \binom{10}{x} 0.1^x 0.9^{(10-x)}$$

(b) Qual é la probabilitá che un pacchetto arrivi corrotto (cioé non tutti i bit sono ricevuti correttamente)?

Soluzione: La probabilitá che il pacchetto ricevuto sia corrotto é

$$1 - P(X = 0) = 1 - \left[\binom{10}{0} 0.1^0 0.9^{(10-0)} \right] = 1 - 0.9^{10} = 0.65132$$

(c) Si supponga che il decimo bit sia un bit di paritá: viene settato a 1 se gli altri 9 bit hanno un numero dispari di 1, a 0 se ne hanno un numero pari. Qual é la probabilitá che un pacchetto corrotto non sia rilevato come tale? (Diciamo che un pacchetto é non corrotto se nessun bit é stato corrotto e il bit di paritá é settato correttamente)

Soluzione:
E' semplice verificare che il bit di paritá indica erroneamente che il pacchetto é "buono" quando un numero pari di bit del pacchetto é stato invertito (includendo eventualmente il bit di paritá medesimo) durante la trasmissione. Questo puó accadere se hanno avuto luogo 2,4,6,8 o 10 errori. Pertanto la probabilitá di non rilevamento $P(\sim R)$ é, per additivitá di tali eventi disgiunti, $P(\sim R) = P(\{X = 2\} \cup \{X = 4\} \cup \{X = 6\} \cup \{X = 8\} \cup \{X = 10\}) = P(\{X = 2\}) + P(\{X = 4\}) + P(\{X = 6\}) + P(\{X = 8\}) + P(\{X = 10\})$. Poiché ciascun evento segue la legge binomiale stabilita precedentemente:

$$P(\sim R) = \binom{10}{2} 0.1^2 0.9^8 + \binom{10}{4} 0.1^4 0.9^6 + \binom{10}{6} 0.1^6 0.9^4 + \binom{10}{8} 0.1^8 0.9^2 + \binom{10}{10} 0.1^{10} 0.9^0 = 0.20501$$

Esercizio 2. (Distribuzione Normale)
Il contenuto reale di birra che un apposito macchinario mette in fusti da 5 litri puó essere considerato come una variabile aleatoria avente una distribuzione normale con una deviazione standard pari a 0.05 litri.

(a) Se solo il 2% dei fusti contiene meno di 5 litri, quale dovrebbe essere il contenuto medio dei fusti?

Soluzione:
Per determinare μ tale che $\Phi(z) = F_X\left(\frac{5-\mu}{\sigma}\right) = 0.02$, si cerca nella tabella della distribuzione normale standard il valore piú vicino a 0.02. Si trova la migliore approssimazione per 0.0202 che corrisponde a $z = -2.05$.
Qualora la tabella a disposizione consideri solo i valori per $z \geq 0$, si faccia uso della proprietá della cumulativa standardizzata $\Phi(-z) = 1 - \Phi(z)$; in tal caso la migliore approssimazione si ha per $1 - \Phi(2.05) = 1 - 0.9798 = 0.0202$.
Pertanto:

$$z = \frac{5 - \mu}{\sigma} = \frac{5 - \mu}{0.05} = -2.05.$$

Risolvendo l'equazione in μ, si trova che $\mu = 5.1$ litri.

Esercizio 3. (Teorema del limite centrale)
Un barattolo di pittura da 1 litro copre in media una superficie di circa 12 metri quadri con una deviazione standard di 0.7 metri quadri.

(a) Qual é la probabilitá che la media campionaria dell'area coperta da un campione di 40 barattoli da 1 litro assuma un valore compreso tra 11.78 e 12.10 metri quadri?

Soluzione:
In breve, il Teorema del limite centrale ci dice che per un numero n grande ($n \to \infty$) di V.A. indipendenti e identicamente distribuite vale

$$\begin{aligned} &\lim_{n\to\infty} P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) = \\ =&\lim_{n\to\infty} P\left(a \leq \frac{\sqrt{n}}{\sigma} \cdot \left(\frac{S_n}{n} - \mu\right) \leq b\right) = \\ =&\Phi(b) - \Phi(a), \end{aligned} \tag{1}$$

dove $S_n = \sum_{i=1}^n X_i$ e $\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}$ é la media campionaria. In buona sostanza, se $\overline{X}$ é la media del campione di taglia n estratto da una popolazione di media μ e varianza σ^2, il teorema garantisce che la variabile ridotta $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ sia una variabile aleatoria la cui funzione di distribuzione tende alla distribuzione normale standard Φ.
Nel caso in esame $\mu = 12$ metri quadri, $\sigma = 0.7$ metri quadri, $n = 40$.
Pertanto, la probabilitá che la media campionaria $\overline{X}$ assuma un valore compreso tra $\overline{X}_a = 11.78$ e $\overline{X}_b = 112.10$ é:

$$P(\overline{X}_a \leq \overline{X} \leq \overline{X}_b) = \Phi\left(\frac{\overline{X}_b - \mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{\overline{X}_a - \mu}{\sigma/\sqrt{n}}\right)$$

dove:

$$a = \left(\frac{\overline{X}_a - \mu}{\sigma/\sqrt{n}}\right) = \left(\frac{11.78 - 12}{0.7/\sqrt{40}}\right) \approx -1.987$$

$$b = \left(\frac{\overline{X}_b - \mu}{\sigma/\sqrt{n}}\right) = \left(\frac{12.10 - 12}{0.7/\sqrt{40}}\right) \approx 0.903$$

Accedendo alla tabella della distribuzione normale standard leggiamo $\Phi(-1.987) = 0.0239$ e $\Phi(0.903) = 0.8238$.
Si ricava dunque che

$$P(\overline{X}_a \leq \overline{X} \leq \overline{X}_b) = \Phi(0.903) - \Phi(-1.987) = 0.791$$