Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 20 febbraio 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.
Un dado non truccato presenta sei facce: 4 verdi e 2 rosse. Si considerino le seguenti sequenze:

`R V R R R`

`R V R R R V`

`V R R R R R`

(a) Vincete 50 euro se, scelta una sequenza ed effettuato un numero di lanci pari alla sua lunghezza, l'esito dei lanci replica la sequenza scelta. Indicate la vostra scelta, motivandola, senza effettuare alcun calcolo probabilistico.

Soluzione:
Assumendo l'indipendenza di ciascun lancio dall'altro, le tre sequenze sono identiche e dunque equiprobabili, nei primi cinque lanci. La seconda e terza sequenza peró comportano un ulteriore lancio e, dunque, la moltiplicazione per un numero minore di uno (per definizione stessa di probabilitá). Dunque la prima sequenza é quella che garantisce maggiormente la vincita.
Una nota curiosa: in un celebre esperimento psicologico (A. Tversky e D. Kahneman, "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment," *Psychological Review*, vol. 90, 1983, pp. 293–315), lo stesso quesito é stato posto a 260 studenti che non avevano nozioni di calcolo delle probabilitá. Il 63% di loro ha scelto la seconda sequenza, mostrando come l'uso della probabilitá che intuitivamente viene praticato nella vita reale non conduca a risultati accurati.

(b) Indicate ora la sequenza che ha maggior probabilitá di garantirvi la vincita derivando il risultato in modo formale (facendo i conti per esteso)

Soluzione:
I dati del problema ci dicono che:

$$P(R) = \frac{2}{6} = \frac{1}{3}$$

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

La soluzione piú veloce é la seguente.
Per indipendenza dei lanci, la probabilitá dei primi cinque é identica: indichiamola con $P(H)$. Pertanto, per la prima sequenza vale

$$P(S_1) = P(H),$$

per la seconda

$$P(S_2) = P(H) \times P(V) = \frac{2}{3}P(H),$$

e per la terza

$$P(S_3) = P(H) \times P(R) = \frac{1}{3}P(H).$$

Da cui:

$$P(S_1) > P(S_2) > P(S_3).$$

Svolgendo, invece, nel dettaglio i conti, per la prima sequenza vale

$$P(S_1) = \frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{2}{243} = 0.0082,$$

per la seconda

$$P(S_2) = \frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} = \frac{4}{729} = 0.0055,$$

e per la terza

$$P(S_3) = \times\frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{2}{729} = 0.0027.$$

Quindi $P(S_1)$ é la probabilitá massima.

Esercizio 2. Un esame consiste in un test a risposte multiple, con cinque possibili risposte per ogni domanda. Mentre sostenete l'esame, vi rendete conto di avere una probabilitá del 75% di formulare la risposta corretta per ogni domanda a cui sapete rispondere. Laddove non sapete rispondere, volete indicare una soluzione con probabilitá di essere corretta pari a $\frac{1}{5}$.

(a) Qual é la probabilitá di rispondere correttamente ad una domanda?

Soluzione:
Definiamo gli eventi

$$A = \text{“risposta corretta”}$$

$$B = \text{“so rispondere”}$$

Si ha che:

$$P(A) = P(A \cap B) + P(A \cap \sim B)$$

dove $\sim B$ é l'evento complementare di B.
Applicando la regola del prodotto e i dati forniti dal problema:

$$P(A \cap B) = P(A \mid B)P(B) = 1 \times 0.75 = 0.75,$$

$$P(A \cap \sim B) = P(A \mid \sim B)P(\sim B) = \frac{1}{5} \times 0.25 = 0.05.$$

Dunque

$$P(A) = P(A \cap B) + P(A \cap \sim B) = 0.75 + 0.05 = 0.8$$

Esercizio 3. Nel paese di Canicattí, il 30% degli elettori si dichiara di sinistra, il 50% di centro, e il 20% di destra. Alle ultime comunali hanno votato il 65% degli elettori di sinistra, l' 82% degli elettori di centro, e il 50% degli elettori di destra.

(a) Scegliendo a caso un elettore che ha votato, qual é la probabilitá che sia di centro?

Soluzione: (Teorema di Bayes)

Definiamo gli eventi:

$$V = \text{“La persona ha votato alle ultime elezioni”}$$
$$S = \text{“La persona e' di sinistra”}$$
$$C = \text{“La persona e' di centro”}$$
$$D = \text{“La persona e' di destra”}$$

La soluzione del problema consiste nel calcolare la probabilitá a posteriori $P(C \mid V)$, ovvero usando Bayes:

$$P(C \mid V) = \frac{P(V \mid C)P(C)}{P(V)}. \tag{1}$$

Sappiamo che:

$$P(S) = 0.3, P(C) = 0.5, P(D) = 0.2$$

$$P(V \mid S) = 0.65, P(V \mid C) = 0.82, P(V \mid D) = 0.5$$

Inoltre:

$$P(V) = P(V \mid S)P(S) + P(V \mid C)P(C) + P(V \mid D)P(D) = 0.705$$

Pertanto sostituendo in (1):

$$P(C \mid V) = \frac{0.82 \times 0.5}{0.705} = 0.5816 \tag{2}$$

Moduli 2 e 3

Esercizio 1.

Una variabile aleatoria X ha densitá $f_X(x) = kx$ per $0 \leq x \leq 3$

(a) Trovare x_1 e x_2 tali che $P(X \leq x_1) = 0.1$ e $P(X \leq x_2) = 0.95$

Soluzione:
Per procedere, trovo la costante k imponendo la normalizzazione

$$\int_0^3 f_X(x)dx = \int_0^3 kxdx = 1$$

da cui ottengo $k = \frac{2}{9}$
Per qualunque $0 \leq x \leq 3$,

$$P(X \leq x) = \frac{2}{9}\int_0^x tdt = \frac{x^2}{9} = F_X(x)$$

.
Pertanto, $\frac{x_1^2}{9} = 0.1$ e $\frac{x_2^2}{9} = 0.95$. Da cui

$$x_1 = \sqrt(0.9) = 0.9487$$

$$x_2 = \sqrt(8.55) = 2.9240$$

(b) Determinare $P(|X - 1.8| < 0.6)$.

Soluzione: Usando la cumulativa $F_X(x)$ calcolata precedentemente:

$$P(|X - 1.8| < 0.6) = P(1.2 < X < 2.4) = \frac{1}{9}(2.4^2 - 1.2^2) = 0.48$$

Esercizio 2.

Si supponga che lo 0.3% di componenti elettronici prodotti da una azienda sia difettoso, dove il numero di componenti difettosi é una variabile aleatoria.

(a) Se i componenti vengono sistemati in scatole da 100, qual é la probabilitá che una scatola contenga un numero x di elementi difettosi?

Soluzione: (Distribuzione binomiale) Il numero di componenti difettosi segue una distribuzione binomiale. Ponendo il parametro $p = 0.003$ (probabilitá di essere difettoso) e $n = 100$

$$P(X = x) = Bin(100, 0.003)$$

(b) Qual é la probabilitá di non trovare elementi difettosi nella scatola?.

Soluzione: Nella fattispecie per i valori di N,p considerati conviene utilizzare un'approssimazione con distribuzione di Poisson, $Poiss(\mu)$, alla Binomiale. La distribuzione di Poisson avrá parametro

$$\mu = np = 100 \times 0.003 = 0.3,$$

ovvero

$$P(X = x) \approx \frac{e^{-0.3}(0.3)^x}{x!},$$

con $x =$ 0,1,2,
Dunque,

$$P(X = 0) \approx 0.7408$$

(c) Si supponga di acquistare 8 scatole. Qual é la distribuzione del numero di scatole con nessun difetto?

Soluzione:
La distribuzione del numero di scatole, diciamo k, con nessun difetto segue di nuovo una distribuzione binomiale

$$P(K = k) = Bin(N, p_S)$$

dove nella fattispecie $N = 8$ e dove p_S indica la probabilitá di una scatola con nessun componente difettoso.
Nel quesito precedente abbiamo stabilito che la probabilitá di una scatola senza componenti difettosi é $P(X = 0) \approx 0.7408$, dunque $p_S \approx 0.7408$. Pertanto:

$$P(K = k) = Bin(8, 0.7408)$$

(d) Qual é il numero atteso di scatole che non contengono elementi difettosi?

Soluzione:

$$E[K] = np = 8 \times 0.7408$$

ESERCIZIO 3. L'altezza di un uomo segue una distribuzione normale di media 178 cm con una deviazione dalla media di 8 cm. L'altezza di una donna ha invece una media di 165 cm con una deviazione pari a 7 cm. .

(a) Qual é, nella stessa popolazione, la proporzione di uomini la cui altezza é superiore a 185 cm?

Soluzione: Sia U la variabile aleatoria che denota l'altezza degli uomini.
Vogliamo determinare $P(U > 185)$, dove la distribuziione é Normale con $\mu = 178, \sigma = 8$. Standardizzando le variabili e usando le tabelle della distribuzione normale

$$P(U > 185) = P(\frac{U - 178}{8} > \frac{185 - 178}{8}) = P(Z < 0.875) \approx 0.19$$

(b) Qual é la proporzione di donne che sono piú alte della metá degli uomini?

Soluzione: Sia D la variabile aleatoria che denota l'altezza delle donne.
Sappiamo che media e mediana nella distribuzione normale coincidono dunque

$$P(U > u) = 0.5$$

quando $u = \mu = 178$
Pertanto, si tratta di determinare $P(D > 178)$:

$$P(D > 178) = P(Z > \frac{178 - 165}{7}) = P(Z > 1.857) \approx 0.032$$

.
Solo il 3.2% delle donne é piú alto di metá degli uomini.