Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 07 novembre 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.

Se gli eventi A e B sono incompatibili, allora $P(A) \leq P(\sim B)$.

(a) Vero o falso?

Soluzione: (Assiomi di probabilitá)

E' vero.

Perché se sono incompatibili allora:

$$A \subseteq \sim B,$$

dove $\sim B = S - B$.

Da cui si deduce, per gli assiomi della probabilitá che $P(A) \leq P(\sim B)$.

Esercizio 2.

Due urne contengono palline bianche e nere in proporzioni diverse. Siano p_1 e p_2 le probabilitá di estrarre una pallina bianca rispettivamente dall'urna 1 e dall'urna 2. Il giocatore vince se estraendo due palline almeno una é bianca. Egli puó scegliere tra due modalitá di estrazione:

A) Sceglie a caso una delle due urne, estrae una pallina, la rimette nell'urna da cui é stata estratta, quindi sceglie di nuovo a caso un'urna ed estrae la seconda pallina.

B) Sceglie a caso una delle due urne, estrae una pallina, la rimette nell'urna da cui é stata estratta, e, sempre dalla stessa urna, estrae una seconda pallina.

(a) Quale tra le due procedure é piú conveniente per la vittoria?

Soluzione: (Probabilitá totali)

Indichiamo con U_i, $i = 1{,}2$ la scelta di una delle due urne.

Definiamo inoltre gli eventi:

$$N_i = \text{“pallina nera alla i-esima estrazione”}$$

$$E = \text{“estrazione di almeno una pallina bianca”}$$

Si ha pertanto che:

$$P(U_i) = 0.5$$

$$P(E) = 1 - P(N_1 \cap N_2)$$

Seguendo la procedura A), le due estrazioni sono statisticamente indipendenti:

$$P(N_1 \cap N_2) = P(N_1)P(N_2) = P(N_1 \mid U_1)P(U_1) + P(N_1 \mid U_2)P(U_2) \times P(N_2 \mid U_1)P(U_1) + P(N_2 \mid U_2)P(U_2) =$$

$$= [(1-p_1)0.5 + (1-p_2)0.5]^2$$

da cui si ricava

$$P_A(E) = 1 - P(N_1 \cap N_2) = 1 - \left[\frac{(1-p_1)}{2} + \frac{(1-p_2)}{2}\right]^2 = p_1 + p_2 - \frac{(p_1+p_2)^2}{4}$$

Seguendo invece la procedura B), la probabilitá di estrarre due palline nere dalla medesima urna (con reimbussolamento della pallina estratta) vale:

$$P(N_1 \cap N_2 \mid U_i) = P(N_1 \mid U_i)P(U_i) = (1-p_i)^2$$

con $i = 1{,}2$. Pertanto, differentemente da A):

$$P(N_1 \cap N_2) = P(N_1 \cap N_2 \mid U_1)P(U_1) + P(N_1 \cap N_2 \mid U_2)P(U_2) = (1-p_1)^2 0.5 + (1-p_2)^2 0.5.$$

Dunque nel caso B)

$$P_B(E) = 1 - P(N_1 \cap N_2) = 1 - \left[\frac{(1-p_1)^2}{2} + \frac{(1-p_2)^2}{2}\right] = p_1 + p_2 - \frac{(p_1^2+p_2^2)}{2}.$$

La differenza fra le due probabilitá é

$$P_A(E) - P_B(E) = -\frac{(p_1+p_2)^2}{4} + \frac{(p_1^2+p_2^2)}{2} = \frac{(p_1-p_2)^2}{4} > 0$$

In conclusione, tra le due strategie é piú conveniente A) essendo $P_A(E) > P_B(E)$.

Esercizio 3.

Due ditte forniscono il medesimo prodotto. Se esso proviene dalla ditta A, la probabilitá che si guasti prima dell?istante t vale $1 - e^{-t}$; se invece proviene dalla ditta B questa probabilitá vale $1 - e^{-2t}$. Il prodotto puó essere acquistato con uguale probabilitá da A o da B, e non é nota la ditta fornitrice. Tuttavia, é stato osservato che il prodotto si guasta in un intervallo di tempo $1 \le t \le 2$.

(a) Determinare la probabilitá che esso sia stato acquistato dalla ditta A.

Soluzione: (Teorema di Bayes)
Definiamo l'evento
$G =$ "guasto in $1 \le t \le 2$"
I dati del problema ci dicono che, a priori, le probabilitá che il prodotto provenga da A o da B valgono

$$P(A) = P(B) = 0.5.$$

La probabilitá di guasto del prodotto A nell'intervallo di tempo $1 \le t \le 2$ vale

$$P(G \mid A) = 1 - e^{-2} - (1 - e^{-1}) = e^{-1} - e^{-2}.$$

Quella del prodotto B nello stesso intervallo é

$$P(G \mid B) = 1 - e^{-2\cdot 2} - (1 - e^{-2\cdot 1}) = e^{-2} - e^{-4}.$$

Applicando la regola di Bayes:

$$P(A \mid G) = \frac{P(G \mid A)P(A)}{P(G \mid A)P(A) + P(G \mid B)P(B)} = \frac{e^{-1} - e^{-2}}{e^{-1} - e^{-2} + e^{-2} - e^{-4}} = \frac{e^2(e-1)}{e^3 - 1} \approx 0.6652$$

Moduli 2 e 3

Esercizio 1.

Una moneta non truccata viene lanciata infinite volte.

(a) Calcolare la probabilitá di osservare 3 volte Testa esattamente e non prima del sesto lancio.

Soluzione: (Distribuzione binomiale) In generale, l'evento di interesse é

$$A = \text{“vengono osservate } k \text{ teste in } n \text{ lanci ma non prima di } n \text{ lanci ”}$$

Tale evento accade se e solo se i due eventi seguenti accadono congiuntamente:

$$B = \text{“vengono osservate } k-1 \text{ teste in } n-1 \text{ lanci ”}$$

$$C = \text{“viene osservata Testa al lancio } n\text{-esimo”}$$

Si ha pertanto che $P(A) = P(B,C)$ e poiché B e C sono indipendenti

$$P(A) = P(B)P(C)$$

Ogni singolo lancio é indipendente dai precedenti, dunque

$$P(C) = P(\text{Testa}) = p = \frac{1}{2}$$

essendo la moneta non truccata.
$P(B)$ ha distribuzione binomiale:

$$P(B) = \binom{n-1}{k-1} p^{k-1} q^{n-1-(k-1)}$$

Pertanto:

$$P(A) = P(B)P(C) = \binom{n-1}{k-1} p^{k-1} q^{n-1-(k-1)} p = \binom{n-1}{k-1} p^{k} q^{n-1-(k-1)} = 0.1562,$$

per $k = 3, n = 6, p = q = \frac{1}{2}$

Esercizio 2.

La variabile aleatoria (v.a.) $X(\omega)$ ha densitá di probabilitá: $\mathrm{f}_X(x) = \begin{cases} k(x-1)^2 & 0 \le x \le 2 \\ 0 & \text{altrimenti} \end{cases}$.

(a) Calcolare la probabilitá che $X(\omega)$ assuma valori in un intorno di raggio $\delta = 0.5$ del suo valor medio.

Soluzione: (Normalizzazone di Variabili aleatorie e media) Anzitutto, determiniamo il valore di k imponendo la condizione di normalizzazione della densitá $f_X(x)$:

$$k \int_0^2 (x-1)^2 dx = 1 \implies k = \frac{3}{2}.$$

Pertanto la densitá correttamente normalizzata é:
$\mathrm{f}_X(x) = \begin{cases} \frac{3}{2}(x-1)^2 & 0 \le x \le 2 \\ 0 & \text{altrimenti} \end{cases}$.
Il problema si risolve calcolando

$$P_X\left(|X - E[X]| < \delta\right) = \cdots \tag{1}$$

Il valore atteso di X é

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

la soluzione (1) si calcola pertanto integrando la densitá nell'intervallo $(1 - 0.5 \le x \le 1 + 0.5)$:

$$P_X\left(|X - E[X]| < \delta\right) = \frac{3}{2} \int_{\frac{1}{2}}^{\frac{3}{2}} (x-1)^2 dx = 3 \int_1^{\frac{3}{2}} (x-1)^2 dx = \frac{1}{8} \tag{2}$$

Esercizio 3.

Un vostro amico asserisce di aver ottenuto una media di 3.25 punti per lancio su 1000 lanci di un dado non truccato.

(a) Qual é la probabilitá che stia mentendo?

Soluzione: (Teorema limite centrale)
Sia X_i la variabile aleatoria (v.a.) che denota l'esito dell'i-simo lancio di dado.
Sappiamo che le v.a. $X_1, X_2, \cdots, X_{1000}$ sono indipendenti e identicamente distribuite con distribuzione uniforme sull'insieme di interi $\{1,2,\cdots,6\}$ da cui é facile ricavare il valore atteso e la deviazione standard

$$E[X_i] = 3.5$$
$$\sigma_{X_i} = \sqrt{E[X_i^2] - E[X_i]^2} \approx 1.7078$$

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 di valore atteso μ e deviazione standard σ vale

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) \tag{3}$$

dove $S_n = \sum_{i=1}^{n} X_i$. Nella fattispecie $n = 1000$ e S_n é il numero di punti totalizzati in 1000 lanci.
Sotto tali ipotesi, la probabilitá che il vostro amico affermi il vero si riduce a calcolare la probabilitá definita nel T.L.C riscritta come

$$P\left(\frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x\right) = \Phi(x) \tag{4}$$

dove $\frac{S_n}{n} = 3.25 = \overline{X}$ é la media campionaria del campione di taglia $n = 1000$ dichiarata dal vostro amico. In buona sostanza il T.L.C. 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. Pertanto,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{3.25 - 3.5}{1.7078/\sqrt{1000}} = -4.629$$

La probabilitá che il vostro amico dichiari il vero é dunque pari a $\Phi(-4.629)$
Anche senza consultare le tabelle e tenendo presente che per la normale standard $\sigma = 1$ e dunque $z = z\sigma$, significa che stiamo valutando la probabilitá di un valore di media dichiarata che é di 4.629 deviazioni standard al di sotto della media $\mu = 3.5$.
Considerando la legge 3σ, é immediatamente evidente che la probabilitá $\Phi(-4.629)$ é molto bassa e pertanto la probabilitá che che il vostro amico dica il falso molto elevata $(1 - \Phi(-4.629))$.
Se poi, per scrupolo, si consultasse la tabella della normale ridotta, si leggerebbe che $\Phi(-4.629) = 1.84 \times 10^{-7}$