Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 21 luglio 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. (Probabilitá condizionata)
Le probabilitá che tre uomini colpiscano un bersaglio sono, rispettivamente $\frac{1}{6}$ $\frac{1}{4}$ e $\frac{1}{3}$.

(a) Determinare la probabilitá che uno solo colpisca il bersaglio

Soluzione: Definiamo i seguenti eventi indipendenti:

- $A =$ `"il primo uomo colpisce il bersaglio"`
- $B =$ `"il secondo uomo colpisce il bersaglio"`
- $C =$ `"il terzo uomo colpisce il bersaglio"`

Il testo del problema ci dice che:

$$P(A) = \frac{1}{6}, P(B) = \frac{1}{4}, P(C) = \frac{1}{3}$$

Denotiamo con $\sim A, \sim B, \sim C$ gli eventi complementari. Ovviamente, $P(\sim A) = 1 - P(A)$, ecc.
L'evento di interesse é

$$T = \texttt{"uno solo colpisce il bersaglio"},$$

ed é definito mediante l'unione di tre eventi mutuamente esclusivi

$$T = \{A \cap \sim B \cap \sim C\} \cup \{\sim A \cap B \cap \sim C\} \cup \{\sim A \cap \sim B \cap C\}.$$

La sua probabilitá é data dunque da:

$$P(T) = P(\{A \cap \sim B \cap \sim C\}) + P(\{\sim A \cap B \cap \sim C\}) + P(\{\sim A \cap \sim B \cap C\})$$

Infine, usando l'indipendenza degli eventi:

$$P(T) = \frac{1}{6} \times (1 - \frac{1}{4}) \times (1 - \frac{1}{3}) + (1 - \frac{1}{6}) \times \frac{1}{4} \times (1 - \frac{1}{3}) + (1 - \frac{1}{6}) \times (1 - \frac{1}{4}) \times \frac{1}{3} = 0.43$$

(b) Determinare la probabilitá che uno solo colpisca il bersaglio e che questi sia il primo uomo.

Soluzione: La probabilitá che l'unico che ha colpito il bersaglio sia il primo uomo é la probabilitá condizionata

$$P(A \mid T) = \frac{P(A \cap T)}{P(T)} = \frac{P(\{A \cap \sim B \cap \sim C\})}{P(T)} = \frac{\frac{1}{6} \times (1 - \frac{1}{4}) \times (1 - \frac{1}{3})}{0.43} = 0.19$$

Esercizio 2. (Legge delle probabilitá totali)
In un primo turno elettorale il polo A ha avuto il 45% dei voti, e il polo B ha vinto con il 55% dei suffragi. Si ripetono le elezioni con i medesimi votanti, e dagli exit-poll risulta che:
- il 10% di coloro che avevano votato A hanno spostato il voto su B;
- il 20% dei vecchi elettori di B hanno votato A.

(a) Chi ha vinto, secondo gli exit-poll il secondo turno?

Soluzione: Definiamo i seguenti eventi:

- $A_1 =$ `"voto per A al primo turno"`
- $B_1 =$ `"voto per B al primo turno"`
- $A_2 =$ `"voto per A al secondo turno"`
- $B_2 =$ `"voto per B al secondo turno"`
- $E =$ `"voto cambiato"`

Dai dati del problema si ricava che:

- $P(A_1) = 0.45$
- $P(B_1) = 0.55$
- $P(E \mid A_1) = 0.10$
- $P(E \mid B_1) = 0.20$

Usando la legge delle probabilitá totali, la probabilitá che gli elettori abbiano votato A al secondo turno é:

$$P(A_2) = P(A_1)\,(1 - P(E \mid A1)) + P(B_1)P(E \mid B_1) = 0.45 \times 0.9 + 0.55 \times 0.20 = 0.515$$

Poiché gli eventi A_2 e B_2 sono mutuamente esclusivi, prendendo in considerazione gli exit-poll, vincerebbe A con il 51.5% contro B che raccoglierebbe il 48.5% dei consensi.

Esercizio 3. (Teorema di Bayes)
In un ufficio le pratiche relative ad una certa procedura amministrativa vengono affidate casualmente a tre impiegati che indicheremo con A, B e C. Statisticamente, la percentuale dei casi in cui la pratica viene completata entro una settimana da ciascun impiegato é riportata nella seguente tabella:

Impiegato	A	B	C
Perc.	40%	80%	30%

(a) Avendo ricevuto una pratica espletata entro una settimana, qual é secondo voi l'impiegato a cui era stata affidata e qual é la probabilitá della vostra conclusione?

Soluzione: Definiamo i seguenti eventi

- $S =$ `"pratica completata entro una settimana"`
- $A =$ `"pratica affidata all'impiegato A"`
- $B =$ `"pratica affidata all'impiegato B"`
- $C =$ `"pratica affidata all'impiegato C"`

L'affidamento pratiche é casuale dunque:

$$P(A) = P(B) = P(C) = \frac{1}{3}$$

Dalla tabella ricaviamo che:

$$P(S \mid A) = 0.4, P(S \mid B) = 0.8, P(S \mid C) = 0.3$$

Per rispondere al quesito si tratta di stabilire qual é la maggiore fra le probabilitá a posteriori

$$P(A \mid S) = \frac{P(S \mid A)P(A)}{P(S)}, P(B \mid S) = \frac{P(S \mid B)P(B)}{P(S)}, P(C \mid S) = \frac{P(S \mid C)P(C)}{P(S)}$$

Il denominatore, cioé la probabilitá che una pratica venga completata entro una settimana indipendentemente da quale impegato l'ha seguita é:

$$P(S) = P(S \mid A)P(A) + P(S \mid B)P(B) + P(S \mid C)P(C) = \frac{1}{3}(0.4 + 0.8 + 0.3) = \frac{1}{3} \times 1.5$$

Sostituendo tutti i dati:

$$P(A \mid S) = \frac{P(S \mid A)P(A)}{P(S)} = 0.267$$

$$P(B \mid S) = \frac{P(S \mid B)P(B)}{P(S)} = 0.533$$

$$P(C \mid S) = \frac{P(S \mid C)P(C)}{P(S)} = 0.2$$

Pertanto, sará stato l'impiegato B a seguire la pratica con probabilitá 0.533.

Moduli 2 e 3

Esercizio 1. (Distribuzione di Poisson, Funzioni generatrici)
Supponiamo di avere un panettone (di volume unitario $V = 1$) in cui ci siano uvette di due tipi, A e B, e che le uvette siano distribuite nel panettone secondo il modello di Poisson, ovvero la variabile aleatoria Poissoniana K_A conta le uvette di tipo A in una porzione di volume v e la variabile aleatoria Poissoniana K_B conta le uvette di tipo B, sempre in una porzione di volume v. Sappiamo che la densitá (numero medio / volume) delle uvette di tipo A é ϱ_A, e la densitá delle uvette di tipo B é ϱ_B.

(a) Calcolare la probabilitá che in una porzione di volume v del panettone ci siano esattamente n uvette.

Soluzione: Il problema si risolve calcolando la probabilitá

$$P(K_A + K_B = n)$$

Per determinare la distribuzione della somma di V.A. $K_A + K_B$ ricorriamo al metodo delle funzioni generatrici. Entrambe le V.A. hanno distribuzione poissoniana. Per una generica distribuzione di Poisson di parametro μ la funzione generatrice corrispondente é

$$G_K(u) = \sum_{k=0}^{\infty} \frac{\mu^k}{k!} e^{-\mu} u^k = e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} u^k = e^{-\mu} \sum_{k=0}^{\infty} \frac{(\mu u)^k}{k!} = e^{-\mu} e^{\mu u} = e^{\mu(u-1)},$$

dove si é fatto uso della serie esponenziale.
Poiché

$$K_A \sim Pois(k_A \mid \mu_A),$$
$$K_B \sim Pois(k_B \mid \mu_B),$$

la distribuzione della VA $K = K_A + K_B$ si puó ottenere mediante le funzioni generatrici:

$$G_K(u) = G_{K_A}(u) G_{K_B}(u) = e^{\mu_{K_A}(u-1)} e^{\mu_{K_B}(u-1)} = e^{(\mu_{K_A}+\mu_{K_B})(u-1)} = e^{\mu(u-1)}$$

dove si vede che la distribuzione $G_K(u)$ é ancora una distribuzione di Poisson, di media $\mu = \mu_{K_A} + \mu_{K_B}$.
(Il risultato é facilmente verificabile derivando k volte la funzione generatrice e ponendo $u = 0$ secondo la definizione

$$P_K(k) = \frac{1}{k!} \frac{d^k}{du^k} e^{\mu(u-1)} |_{u=0} = \frac{1}{k!} \mu^k e^{-\mu}$$

)
Dai dati del problema:

$$\mu_{K_A} = \varrho_A v, \mu_{K_B} = \varrho_B v.$$

Pertanto:

$$P(K_A + K_B = n) = \frac{((\varrho_A + \varrho_B)v)^n}{n!} e^{-(\varrho_A+\varrho_B)v}$$

Esercizio 2. (Distribuzione Normale, uso di tabelle)
La durata in giorni di una gravidanza stribuita come una normale di media 270 giorni e deviazione standard 10 giorni. Una signora un po' distratta cerca di capire chi sia il padre di suo figlio. Un suo partner é stato all'estero dal 290 al 240-esimo giorno antecedente la nascita del bambino.

(a) Se quest'uomo é veramente il padre del bambino qual é la probabilitá che la signora abbia avuto una gravidanza cosí corta o cosí lunga?

Soluzione: Definiamo:

- $T =$ `durata della gravidanza`

Il concepimento prima dei 290 giorni oppure dopo il 240-esimo giorno dalla nascita puó essere definito come:

- $C = (T > 290) \cup (T < 240)$

La probabilitá di interesse é dunque:

$$P(C) = P((T > 290) \cup (T < 240)) = P(T > 290) + P(T < 240)$$

Standardizzando le variabili:

$$P(C) = P(\frac{T-270}{10} > \frac{290-270}{10}) + P(\frac{T-270}{10} < \frac{240-270}{10}) = P(\frac{T-270}{10} > 2) + P(\frac{T-270}{10} < -3),$$

ovvero, usando le proprietá della CDF normale standard:

$$P(C) = 1 - P(\frac{T-270}{10} \leq 2) + P(\frac{T-270}{10} < -3) = 1 - \Phi(2) + \Phi(-3) = 1 - \Phi(2) + 1 - \Phi(3).$$

Accedendo alla tabella della Normale standard per $z = 2$ e $z = 3$ e leggendo i valori di $\Phi(2)$ e $\Phi(3)$:

$$P(C) = 0.0241$$

Una probabilitá invero non molto elevata.

Esercizio 3. (Distribuzioni Esponenziale negativa e Gaussiana)
Due ipotesi sono in competizione per la spiegazione di un fenomeno sperimentale.
La prima assume che dato un input $T = n$ (n, numero intero, per semplicitá), l'output Y di un sistema dovrebbe essere una variabile aleatoria che segue una legge esponenziale negativa con parametro $\lambda = \frac{1}{n}$. L'ipotesi alternativa postula che a paritá di input, Y segue una legge Normale con parametri $\mu = n, \sigma^2 = n^2$.

(a) Come si differenziano le due ipotesi nel predire la probabilitá di un output tale che $Y > 2n$ per $T = n$?

Soluzione: Per un V.A. con legge esponenziale di parametro $\lambda = \frac{1}{n}$,

$$P(Y > 2n) = S(2n) = e^{-\lambda 2n} = e^{-\frac{1}{n}2n} = e^{-2} = 0.13534$$

Per una V.A. con legge Normale, riscrivendo $P(Y > 2n)$ in forma standard,

$$P\left(\frac{Y-n}{n} > \frac{2n-n}{n}\right) = P(Z > 1) = 1 - \Phi(1)$$

Usando la tabella della Normale standard si ha che

$$P(Y > 2n) = 0.15866$$

I due valori si differenziano ma non troppo.

(b) Viene effettuato un esperimento per $T = 10$ e si osserva a posteriori che $Y \in (9{,}12)$. Assumendo una probabilitá a priori uniforme sugli input $P(T = n) = \frac{1}{10}$, $n = 1{,}2{,}3 \cdots 10$, quale delle due ipotesi é piú probabile ovvero spiega meglio l'osservazione effettuata?

Soluzione: Si tratta in buona sostanza di calcolare la probabilitá a a posteriori $P(T = 10 \mid 9 < Y < 12)$ per entrambe le ipotesi usando la regola di Bayes:

$$P(T = 10 \mid 9 < Y < 12) = \frac{P(9 < Y < 12 \mid T = 10)P(T = 10)}{P(9 < Y < 12)}$$

Per l'ipotesi esponenziale la verosimiglianza vale:

$$P(9 < Y < 12 \mid T = n) = F_{T=n}(12) - F_{T=n}(9) = 1 - e^{-\lambda 12} - 1 + e^{-\lambda 9} = e^{-\frac{1}{n}9} - e^{-\frac{1}{n}12}$$

Il termine a denominatore (l'evidenza), ricordando che rappresenta la probabilitá totale

$$P(9 < Y < 12) = \sum_{n=1}^{10} P(9 < Y < 12 \mid T = n)P(T = n),$$

vale:

$$P(9 < Y < 12) = \left(F_T(12) - F_{T(9)}\right)\frac{1}{10} = \sum_{n=1}^{10} \frac{1}{10}\left(e^{-\frac{1}{n}9} - e^{-\frac{1}{n}12}\right).$$

Pertanto,

$$P(T = 10 \mid 9 < Y < 12) = \frac{\left(e^{-\frac{1}{n}9} - e^{-\frac{1}{n}12}\right)\frac{1}{10}}{\sum_{n=1}^{10}\frac{1}{10}\left(e^{-\frac{1}{n}9} - e^{-\frac{1}{n}12}\right)} \approx 0.15828.$$

Per l'ipotesi Normale si procede esattamente esattamente nello stesso modo, scrivendo la verosimiglianza forma standard:

$$P\left(9 < Y < 12 \mid T = n\right) = P\left(\frac{9-n}{n} < \frac{Y-n}{n} < \frac{12-n}{n} \mid T = n\right) = \Phi\left(\frac{12}{n} - 1\right) - \Phi\left(\frac{9}{n} - 1\right),$$

e dunque

$$P(9 < Y < 12) = \sum_{n=1}^{10} \frac{1}{10}\left(\Phi\left(\frac{12}{n} - 1\right) - \Phi\left(\frac{9}{n} - 1\right)\right).$$

Per l'ipotesi Gaussiana abbiamo infine:

$$P(T = 10 \mid 9 < Y < 12) = \frac{\Phi(0.2) - \Phi(-0.1)}{\sum_{n=1}^{10}\left(\Phi\left(\frac{12}{n} - 1\right) - \Phi\left(\frac{9}{n} - 1\right)\right)} \approx 0.1$$

Nel calcolare il denominatore, per velocizzare, notiamo che $z_n = \frac{12}{n} - 1$ da' origine alla sequenza

$$z_n = 11.0000, 5.0000, 3.0000, 2.0000, 1.4000, 1.0000, 0.7143, 0.5000, 0.3333, 0.2000,$$

mentre per il termine $z_n = \frac{9}{n} - 1$

$$z_n = 8.0000, 3.5000, 2.0000, 1.2500, 0.8000, 0.5000, 0.2857, 0.1250, 0, -0.1000.$$

Eliminiamo i termini uguali e utilizziamo l'approssimazione $P(Z < 3.5) \approx 1$. Per i termini restanti, leggiamo $\Phi(z_n)$ in tabella.
Ne concludiamo che l'ipotesi esponenziale é da preferirsi. Notiamo anche che l'ipotesi Gaussiana produce un valore a posteriori ($\approx 1\frac{}{10}$) che non si differenzia dalla probabilitá a priori $P(T = n) = \frac{1}{10}$