Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 23 giugno 2016	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta è di 2 ore e 30 minuti. Durante la prova è 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 più passaggi di calcolo, non sarà preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o più 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 sarà 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)
Un cappello contiene tre carte: una carta é nera su entrambi i lati; una carta é bianca su entrambi i lati; una carta é nera su un lato e bianca sull'altro. Le carte vengono mescolate nel cappello, poi una viene estratta a caso e appoggiata su di un tavolo.

(a) Se il lato visibile della carta é nero, qual é la probabilitá che l'altro lato sia bianco?

Soluzione: Uno studente brillante potrebbe risolvere in pochi secondi il problema ragionando come segue: la carta di interesse deve essere o la nero / nero o la nero / bianco: queste hanno pari probabilitá, dunque la probabilitá che, osservato il nero, l'altro lato sia bianco é $\frac{1}{2}$.
Lo studente, successivamente, per verificare la sua conclusione, effettua una simulazione del gioco delle carte (si veda lo script Matlab carte.m allegato) ma ottiene il risultato mostrato in Figura 1.

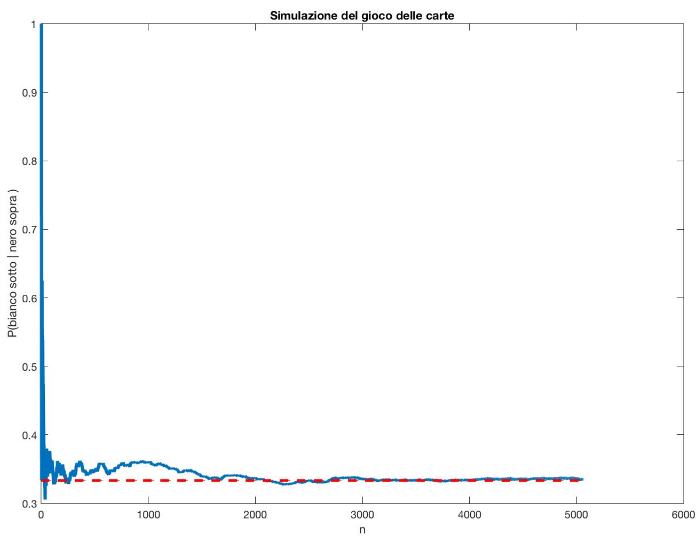

Figure 1: Simulazione del gioco delle tre carte: il valore di $P($“bianco sotto” $\mid$ “nero sopra”$)$ valutato in termini di frequenza relativa, per un numero di prove n crescente, tende al valore di $0.333\cdots \approx \frac{1}{3}$
.

Il valore dell'approssimazione frequentistica di P("bianco sotto" | "nero sopra") tende a $\frac{1}{3}$.
Per capire bene come si arriva a questo risultato, mettiamo da parte l'intuizione e definiamo con precisione lo spazio campionario e gli eventi di interesse.
Identifichiamo i lati delle carte

- N_1 e N_2 per la carta nero / nero
- B_1 e B_2 per la carta bianco / bianco
- N_3 e B_3 per la carta nero / bianco

In questo caso lo spazio campionario S é

$$S = \{N_1, N_2, N_3, B_1, B_2, B_3\}$$

L'evento $N =$ "nero sopra" é definibile come $N = \{N_1, N_2, N_3\}$.
L'evento $B =$ "bianco sotto" é definibile come $B = \{N_3, B_1, B_2\}$
Pertanto:

$$P(B \mid N) = \frac{P(B \cap N)}{P(N)}$$

Poiché #("bianco sotto" $\cap$ "nero sopra") = 1 e #("nero sopra") = 3 :

$$P(B \mid N) = \frac{1}{3}.$$

Esercizio 2. (Legge delle probabilitá totali)
Ci sono due squadre, 1 e 2, di rigoristi. La squadra i-sima ha $3i$ componenti. Per ogni rigore battuto, un rigorista della squadra i ha una probabilitá $\frac{1}{i+1}$ di fare goal, indipendentemente dai rigori battuti precedentemente.

(a) Un rigorista viene selezionato casualmente fra tutti i rigoristi disponibili. Sia G l'evento che venga segnato un goal. Calcolare la probabilitá $P(G)$ che un goal sia realizzato, quale che sia la squadra di provenienza del rigorista.

Soluzione: In prima battuta, la probabilitá dell'evento G che un rigorista segni un goal é condizionata dalla squadra a cui appartiene: $P(G \mid S_i) = \frac{1}{i+1}$. Pertanto:

$$P(G \mid S_1) = \frac{1}{2}$$
$$P(G \mid S_2) = \frac{1}{3}$$

Ciascuna squadra ha $3i$ componenti: dunque 3 per la squadra 1 e 6 della squadra 2. Indicando con S_i l'evento che sia stato selezionato in maniera casuale un rigorista della squadra i-esima, le probabilitá a priori sono

$$P(S_1) = \frac{3}{9} = \frac{1}{3}$$
$$P(S_2) = \frac{6}{9} = \frac{2}{3}$$

Usando la Legge delle Probabilitá totali:

$$P(G) = P(G \mid S_1)P(S_1) + P(G \mid S_2)P(S_2) = \frac{1}{2} \times \frac{1}{3} + \frac{1}{3} \times \frac{2}{3} = \frac{7}{18} = 0.388$$

(b) Vengono scelti a caso due rigoristi. Per $j = 1,2$, sia G_j l'evento che il rigorista j faccia goal. Trovare la probabilitá che il rigorista 1 e il rigorista 2 segnino entrambi un goal (*Suggerimento:* calcolare, mediante un chance tree, la probabilitá dell'evento S_{kl} che il primo rigorista sia stato selezionato dalla squadra k e il secondo rigorista dalla squadra l, con $k,l = 1,2$)

Soluzione: Per risolvere questa parte dl problema ricaviamo l'albero di probabilitá per due "estrazioni" successive e casuali di un rigorista dalle squadre 1 e 2 rappresentato in Figura 2,
Il diagramma ci dice che:

$$P(S_{11}) = \frac{1}{12}, P(S_{12}) = \frac{1}{4}, P(S_{21}) = \frac{1}{4}, P(S_{22}) = \frac{5}{12}$$

La probabilitá dell'evento congiunto $\{G_1, G_2\}$ é condizionata da quali squadre sono stati selezionati i rigoristi 1 e 2, ovvero dall'evento S_{kl} le cui probabilitá sono state derivate con il diagramma ad albero. Essendo l'esito di un rigore indipendente dall'altro: $P(G_1, G_2 \mid S_{kl}) = P(G_1 \mid S_{kl})P(G_2 \mid S_{kl})$. Pertanto:

$$P(G_1, G_2 \mid S_{11}) = \left(\frac{1}{2}\right)^2$$

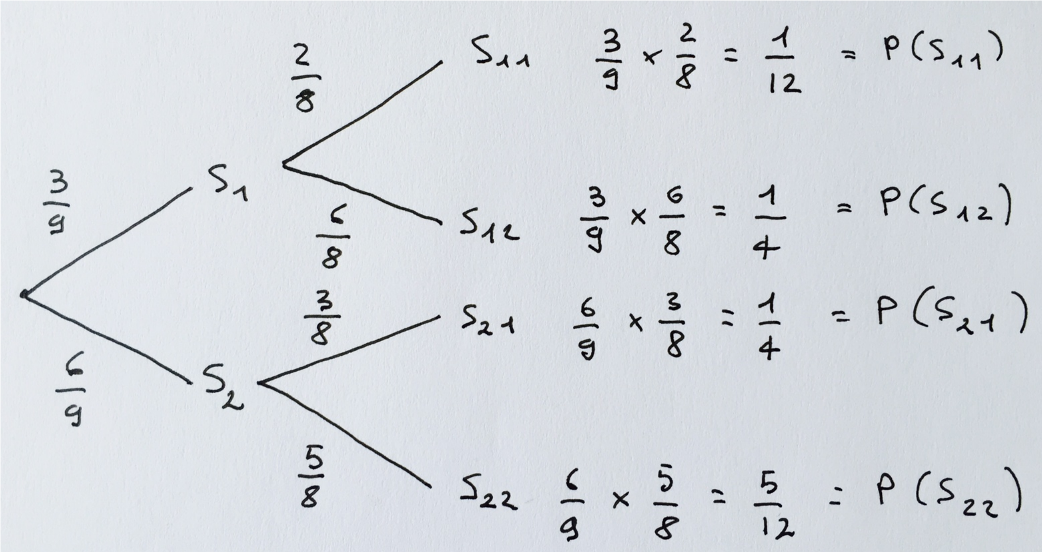

Figure 2: Chance tree per due "estrazioni" successive e casuali di un rigorista dalle squadre 1 e 2: S_k é l'evento che sia stato selezionato in maniera casuale un rigorista della squadra i-esima, S_{kl} che il primo rigorista sia stato selezionato dalla squadra k e il secondo rigorista dalla squadra l

$$P(G_1, G_2 \mid S_{12}) = \frac{1}{2} \times \frac{1}{3}$$
$$P(G_1, G_2 \mid S_{21}) = \frac{1}{3} \times \frac{1}{2}$$
$$P(G_1, G_2 \mid S_{22}) = \left(\frac{1}{3}\right)^2$$

Usando di nuovo la Legge delle Probabilitá totali:

$$P(G_1, G_2) = P(G_1, G_2 \mid S_{11})P(S_{11}) + P(G_1, G_2 \mid S_{12})P(S_{12}) + P(G_1, G_2 \mid S_{21})P(S_{21}) + P(G_1, G_2 \mid S_{22})P(S_{22}) = \frac{15}{96} = 0.1562$$

(c) Verificare se G_1 e G_2 siano eventi indipendenti. Commentare brevemente il risultato ottenuto.

Soluzione: Possiamo verificare la condizione di indipendenza stocastica:

$$P(G_1, G_2) = P(G_1)P(G_2)$$

E' evidente che $P(G_1) = P(G) = \frac{7}{18}$ e che per simmetria $P(G_1) = P(G_2)$, giacché lo scambio dei due rigoristi non muta nulla in termini di esito (se non si nvinti, usare ancora le probabilitá totali per calcolare $P(G_i) = P(G_i \mid S_{11})P(S_{11}) + P(G_i \mid S_{12})P(S_{12}) + P(G_i \mid S_{21})P(S_{21}) + P(G_i \mid S_{22})P(S_{22}) = \frac{7}{18}$).
Si ha allora che:

$$P(G_1, G_2) = 0.1562 \neq 0.1512 = (0.388)^2 = P(G_1)P(G_2)$$

dunque non sono indipendenti. Le due probabiliá ricavate sono vicine ma non esattamente uguali. G_1 e G_2 sono dipendenti perché se il primo rigore viene segnato, vi é maggior probabilit che abbia tirato un rigorista del gruppo 1. Questo rende piú probabile che il secondo giocatore venga selezionato dal gruppo 2, evento che riduce la probabilitá di segnare al secondo tiro.

Esercizio 3. (Teorema di Bayes)

Un robot autonomo controlla il livello di corrosione interna delle tubature dei sistemi di raffreddamento di una centrale nucleare. La corrosione non puó essere osservata direttamente, ma il robot effettua un test che puó fornisce indizi sulla specifica corrosione. Il test non é infallibile: nel 70% dei casi individua una corrosione quando é presente, ma puó anche sbagliare indicando una corrosione inesistente nel 20% dei casi. Risolvere il problema del robot a cui un operatore remoto, sotto l'ipotesi che il 10% delle tubature siano corrose, chiede di:

(a) Comunicargli la probabilitá che una parte delle tubature abbia delle corrosioni interne dopo che il test ha dato esito positivo (presenza di corrosioni). Determinare tale probabilitá.

Soluzione: Definiamo gli eventi $C =$ "la tubatura é corrosa" e $\overline{C} =$ "il test ha identificato la tubatura come corrosa". I dati del problema ci dicono che, a priori,

$$P(C) = 0.1$$

e dunque $P(\sim C) = 1 - 0.1 = 0.9$, dove $\sim C =$ "la tubatura non é corrosa". Inoltre

$$P(\overline{C} \mid C) = 0.7$$
$$P(\overline{C} \mid\sim C) = 0.2$$

Applicando la regola di Bayes:

$$P(C \mid \overline{C}) = \frac{P(\overline{C} \mid C)P(C)}{P(\overline{C})}$$

dove $P(\overline{C}) = P(\overline{C} \mid C)P(C) + P(\overline{C} \mid\sim C)P(\sim C) = 0.25$. Quindi

$$P(C \mid \overline{C}) = \frac{0.7 \times 0.1}{0.25} = 0.28$$

(b) Comunicargli la probabilitá che una parte delle tubature abbia delle corrosioni dopo che il test ha dato esito negativo (nessuna corrosione). Determinare tale probabilitá.

Soluzione: Si ripeta il procedimento precedente per il caso $P(C \mid\sim \overline{C})$

Moduli 2 e 3

Esercizio 1. (Funzione di densitá e di ripartizione)

Data la funzione: $\mathrm{f}_X(x) = \begin{cases} \frac{x}{2} - 1 & 2 \leq x < k \\ 0 & \text{altrimenti} \end{cases}$.

(a) Si determini il valore di k che assicura che $f_X(x)$ rappresenta una funzione di densitá.

Soluzione: Imponiamo la condizione di normalizzazione della densitá $f_X(x)$:

$$\int_2^k \left(\frac{x}{2} - 1\right) dx = 1 \implies \left[\frac{x^2}{4}\right]_2^k - (k-2) = 1 \implies k^2 - 4k = 0$$

Risolvendo l'equazione, vi sono due possibili radici $k_1 = 0$ e $k_2 = 4$. Poiché $k_1 < 2 \leq x < k$, é da scartare, dunque $k = k_2 = 4$. Pertanto la densitá correttamente normalizzata é:

$$f_X(x) = \frac{x}{2} - 1 \text{ con } 2 \leq x < 4$$

(b) Si individui la corrispondente funzione di ripartizione $F_X(x)$

Soluzione: Per definizione $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$. Nella fattispecie:

$$F_X(x) = \int_2^x \left(\frac{t}{2} - 1\right) dt$$

Integrando come fatto prima si ottiene:

$$F_X(x) = \frac{x^2}{4} - x + 1$$

per $2 \leq x < 4$

(c) Si individui la mediana della variabile aleatoria X descritta dalla densitá $f_X(x)$

Soluzione: Per definizione la mediana x_{med} é il valore per cui

$$F_X(x) = \frac{1}{2} \implies \frac{x^2}{4} - x + 1 = \frac{1}{2}$$

Risolvendo l'equazione si trovano le soluzioni $x_1 = 2 + \sqrt{2}$ e $x_2 = 2 - \sqrt{2}\sqrt{2}$. Dunque $x_{med} = x_1$

Esercizio 2. (Distribuzioni di Poisson ed Esponenziale negativa)

Il numero di automobili che attraversano un particolare incrocio stradale in un'ora é mediamente pari a 30.

(a) Determinare la probabilitá che in un intervallo di tempo di cinque minuti nessuna automobile attraversi l'incrocio in questione.

Soluzione: Sia X la V.A. discreta che indica il numero di veicoli che passa lungo l'incrocio in 5 min. X segue una legge di Poisson, $X \sim Pois(\mu)$. Il numero medio di automobili che passa in 5 min. é $\frac{30}{12} = 2.5$. Dunque $X \sim Pois(2.5)$. Pertanto:

$$P(X=0) = \frac{\mu^0}{0!}e^{-\mu} = e^{-2.5} = 0.082$$

(b) Qual é la probabilitá che in dieci minuti almeno due automobili passino lungo quel tratto di strada?

Soluzione: Se X denota il numero di veicoli che passa lungo l'incrocio in 5 min, indichiamo con $Y = 2X$ la V.A. discreta che indica il numero di veicoli che passa lungo l'incrocio in 10 min. Allora:

$$E[Y] = E[2X] = 2E[X] = 2\mu = 5$$

Dunque, $Y \sim Pois(5)$. La risposta alla domanda consiste nel calcolare

$$P(Y \geq 2) = 1 - P(Y < 2) = 1 - (P(Y=0) + P(Y=1)) = 1 - \left(e^{-5} + \frac{5^1}{1!}e^{-5}\right) = 1 - 6e^{-5} = 0.9595$$

(c) Assumendo che la variabile aleatoria T, rappresentante il tempo (sempre in minuti) trascorso tra il passaggio di un'auto e di quella successiva, si distribuisca con legge esponenziale negativa, qual é la probabilitá che tra il passaggio di un'auto e la successiva trascorra piú di un minuto?

Soluzione: Poiché $T \sim Exp(\lambda)$, i dati del problema ci dicono che il rate di passaggio é

$$\lambda = \frac{30}{60}\frac{\text{(auto)}}{\text{(min.)}}.$$

Dunque la V.A. T ha funzione di densitá $f_T(t) = 0.5e^{-0.5t}$, $t \geq 0$. La probabilitá che il tempo di attesa tra il passaggio di un un'auto e la successiva sia > 1 min. é $P(T > 1)$ che si puó ottenere dal calcolo diretto $P(T > 1) = \int_1^\infty f_T(t)dt$, oppure usando la funzione di sopravvivenza $S_T(t) = e^{-\lambda t}$:

$$P(T > 1) = 1 - F_T(1) = S_T(1) = e^{-0.5 \cdot 1} = 0.6065$$

Esercizio 3. (Distribuzione Normale, uso di tabelle)

Una linea di produzione fabbrica resistori da 1000 ohm (Ω). La fabbrica adotta una tolleranza del 10% (ovvero sono scartati i resistori con resistenza $R < 900\ \Omega$ e $R > 1100\ \Omega$).

(a) Sotto l'ipotesi che la resistenza R di un resistore sia una variabile aleatoria distribuita con legge normale di media $1000\ \Omega$ e varianza 2500, si calcoli la probabilitá che un resistore preso a caso venga scartato

Soluzione: Sia S l'evento in cui un resistore viene scartato. Allora:

$$S = \{R < 900\} \cup \{R > 1100\}$$

Poiché $\{R < 900\} \cap \{R > 1100\} = \emptyset$, si ha che

$$P(S) = P(\{R < 900\}) + P(\{R > 1100\}) = F_X(900) + (1 - F_X(1100)) = \Phi\left(\frac{900-\mu}{\sigma}\right) + \left(1 - \Phi\left(\frac{1100-\mu}{\sigma}\right)\right)$$

Poiché $\mu = 1000$, e $\sigma^2 = 2500 \implies \sigma = 50$, le variabili normalizzate sono

$$\frac{900-1000}{50} = -2$$

$$\frac{1100-1000}{50} = 2$$

.

Sfruttando la proprietá della cumulativa standardizzata $\Phi(-z) = 1 - \Phi(z)$, la precedente diventa:

$$P(S) = 2\,(1 - \Phi(2))$$

.

Accedendo alla tabella della Normale standard per $z = 2$, $\Phi(2) = 0.9772$ da cui $P(S) = 0.0455$