Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 26 gennaio 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.
Dovete partecipare ad un torneo di scacchi in cui giocate contro altri tre giocatori (una sola partita con ciascuno). Conoscete le probabilitá di vincere con ognuno di loro e potete scegliere l'ordine in cui incontrarli. Si vince il torneo se si vincono due partite consecutive.

(a) Volendo massimizzare la probabilitá di vittoria, si mostri che l'ordinamento ottimale é quello che vi fa incontrare alla seconda partita il giocatore piú debole dei tre, mentre l'ordine in cui incontrate i restanti due giocatori non ha rilevanza.

Soluzione:
Denotiamo con

$$P(\texttt{"vittoria contro il giocatore incontrato alla partita i -esima"}) = P(i) = p_i$$

con $i = 1,2,3$.
La descrizione del problema ci dice che si vince la partita se si gioca la seconda partita con il giocatore piú debole - dunque con $p_2 \geq p_1, p_2 \geq p_3$ - e, **congiuntamente**, si vince con l'uno **oppure** l'altro giocatore incontrati alla partita 1 o 3. Quindi, formalmente, la probabilitá $P(V)$ di vincere il torneo é

$$\begin{aligned} P(V) = P(2 \cap (1 \cup 3)) = P(2)P(1 \cup 3) = P(2)\left[P(1) + P(3) - P(1) \cap P(3)\right] = \\ = p_2(p_1 + p_3 - p_1 p_3). \end{aligned}$$

L'ordinamento é ottimale se tale probabilitá non é inferiore alle probabilitá calcolate utilizzando i due ordinamenti alternativi $P(1 \cap (2 \cup 3)) = p_1(p_2 + p_3 - p_2 p_3)$ e $P(3 \cap (2 \cup 1)) = p_3(p_2 + p_1 - p_2 p_1)$, ovvero:

$$p_2(p_1 + p_3 - p_1 p_3) \geq p_1(p_2 + p_3 - p_2 p_3), \quad (1)$$
$$p_2(p_1 + p_3 - p_1 p_3) \geq p_3(p_2 + p_1 - p_2 p_1). \quad (2)$$

Infatti, la prima disuguaglianza é equivalente alla condizione

$$p_2 \geq p_1,$$

mentre la seconda disuguaglianza si riduce a

$$p_2 \geq p_3.$$

Le due condizioni (1) e (2) ci dicono che la probabilitá di vincere alla seconda partita deve essere massima, ovvero: per massimizzare la probabilitá di vincere il torneo il secondo giocatore da incontrare deve essere il piú debole dei tre.

Esercizio 2. Un lotto di cento componenti elettronici viene ispezionato da un compratore scegliendo quattro componenti a caso: se uno dei quattro componenti risulta essere difettoso, l'intero lotto viene scartato.

(a) Qual é la probabilitá che il lotto venga acquistato dal compratore se contiene cinque componenti difettosi?

Soluzione:
Definiamo gli eventi

$$A = \text{“il lotto viene acquistato”}$$

$$A_i = \text{“il componente i-simo non é difettoso”}$$

con $i = 1, 2, 3, 4$.
E' evidente che:

$$A = A_1 \cap A_2 \cap A_3 \cap A_4$$

Applicando banalmente la regola del prodotto

$$P(A) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_2, A_1)P(A_4 \mid A_3, A_2, A_1) = \frac{95}{100} \times \frac{94}{99} \times \frac{93}{98} \times \frac{92}{97} = 0.812$$

Esercizio 3. Marta cerca un documento che sa di avere archiviato in uno dei suoi raccoglitori. Presume di averlo lasciato nel raccoglitore j-esimo con probabilitá $p_j > 0$.
Purtroppo ciascun raccoglitore é cosí disordinato che anche se Marta decidesse correttamente di cercare il documento nel raccoglitore i-esimo che lo contiene, la probabilitá di trovarlo effettivamente nel raccoglitore sarebbe solo pari a r_i.
Alla fine Marta sceglie un particolare raccoglitore, diciamo i, ma il documento non viene trovato.

(a) Subordinatamente all'esito della ricerca, determinare la probabilitá che il documento fosse nel raccoglitore j nel caso $j \neq i$ e la probabilitá che il documento fosse in j nel caso $j = i$

Soluzione: (Teorema di Bayes)
Definiamo l'evento E che sintetizza l'esito della ricerca:

$$E = \text{“Il documento non viene trovato nel raccoglitore } i\text{”}$$

I dati del problema ci dicono che il documento é nel raccoglitore i con probabilitá p_i e che puó essere trovato con probabilitá r_i. Pertanto, la probabilitá dell'evento complementare

$$\sim E = \text{“Il documento viene trovato nel raccoglitore } i\text{”}$$

é pari a

$$P(\sim E) = p_i r_i,$$

ovvero:

$$P(E) = 1 - P(\sim E) = 1 - p_i r_i. \tag{3}$$

Denotiamo ora con B l'evento:

$$B = \text{“Il documento é nel raccoglitore } j\text{”}.$$

Vogliamo determinare nei due casi $j \neq i$ e $j = i$ la probabilitá condizionata

$$P(\text{ “Il documento é nel raccoglitore } j\text{”} \mid \text{ “Il documento non viene trovato nel raccoglitore } i\text{”}) = P(B \mid E),$$

ovvero vogliamo calcolare, usando Bayes, la probabilitá a posteriori

$$P(B \mid E) = \frac{P(E \mid B)P(B)}{P(E)} \tag{4}$$

1. $j \neq i$:
   In questo caso $E \cap B = B$. Se sappiamo che il documento é in j con $j \neq i$, la probabilitá che non venga trovato in i é $P(E \mid B) = 1$ (l'evento é certo). Sostituendo in (4) e usando la (3):

$$P(B \mid E) = \frac{1 \times P(B)}{P(E)} = \frac{p_j}{1 - p_i r_i}. \tag{5}$$

2. $j = i$:
   In questo caso la probabilitá

   $$P(E \mid B) = P(\text{ “Il documento non viene trovato nel raccoglitore } i\text{” } \mid \text{ “Il documento é nel raccoglitore } i\text{”}) = 1 - r_i$$

   Di nuovo, sostituendo in (4) e usando la (3):

   $$P(B \mid E) = \frac{(1 - r_i)P(B)}{P(E)} = \frac{(1 - r_i)p_j}{1 - p_i r_i}. \tag{6}$$

Esercizio 1.
Un sistema di comunicazione consiste di un buffer che ospita i pacchetti provenienti dalla sorgente, e da una linea di comunicazione che recupera i pacchetti dal buffer e li trasmette a un ricevitore. Il sistema opera a coppie di *time-slot*. Nel primo slot il sistema mette nel buffer un certo numero di pacchetti generati dalla sorgente secondo una distribuzione di Poisson di parametro μ. La capacitá del buffer (numero massimo di pacchetti) é pari all'intero b_{max}; i pacchetti che arrivano a buffer pieno vengono scartati. Nel secondo slot temporale il sistema trasmette c pacchetti al ricevitore, dove c é un intero $0 < c < b_{max}$, oppure trasmette quelli effettivamente presenti se in numero inferiore a c.

(a) Assumendo che all'inizio del primo slot il buffer sia vuoto, trovare la PMF del numero di pacchetti bufferizzati alla fine del primo slot e la PMF alla fine del secondo slot.

Soluzione: Sia X la V.A. discreta che indica il numero di pacchetti bufferizzati alla fine del primo slot temporale Δt_1. La Figura 1 mostra il suo intervalli di variazione relativamente ai parametri del problema c, b_{max} Per determinare la PMF di X, P_X é sufficiente determinare: 1) la probabilitá di avere bufferizzato $k = 0,1,2,\cdots,b_{max}-1$ pacchetti; 2) la probabilitá di avere il buffer pieno che accade quando in Δt_1 vengono generati alla sorgente $k \geq b_{max}$ pacchetti.

1. $k = 0,1,2,\cdots,b_{max}-1$:
   La probabilitá di avere esattamente k pacchetti bufferizzati $P_X(X = k)$ é uguale alla probabilitá che k pacchetti siano stati generati alla sorgente durante Δt_1: ció implica, per quanto ci dice la specifica del problema, che X segua una legge di Poisson, $X \sim Pois(\mu)$ e dunque:
   $$P_X(X = k) = \frac{\mu^k}{k!}e^{-\mu} \tag{7}$$
   applicabile per $k = 0,1,2,\cdots,b_{max}-1$.
2. $k \geq b_{max}$:
   La probabilitá di avere il buffer pieno, ovvero che $X = b_{max}$ é la probabilitá che alla sorgente siano stati generati almeno b_{max} pacchetti, ovvero
   $$P_X(X = b_{max}) = \sum_{k=b_{max}}^{\infty} \frac{\mu^k}{k!}e^{-\mu} = 1 - \sum_{k=0}^{b_{max}-1} \frac{\mu^k}{k!}e^{-\mu}$$

Si noti come $P_X(X = 0) + P_X(X = 1) + \cdots + P_X(X = b_{max}-1) + P_X(X = b_{max}) = 1$, dunque P_X é una PMF correttamente normalizzata.
Chiamiamo ora Y la V.A. discreta che indica il numero di pacchetti bufferizzati alla fine del secondo slot temporale Δt_2. Il suo valore é determinabile come

$$Y = \text{ pacchetti bufferizzati durante } \Delta t_1 - \text{ pacchetti trasmessi durante } \Delta t_2 = X - min\{X,c\}.$$

Infatti se $X \leq c$, allora $min\{X,c\} = X$ e $Y = X - X = 0$. Se $X > c$, $min\{X,c\} = c$, quindi $Y = X - c$.
Per determinare la PMF di Y consideriamo i seguenti casi

1. $Y = 0$: la probabilitá che il buffer sia stato svuotato é uguale alla probabilitá che vi fossero $X \leq c$ pacchetti bufferizzati in Δt_1:
   $$P_Y(Y = 0) = P_X(X \leq c) = \sum_{k=0}^{c} P_X(X = k) = \sum_{k=0}^{c} \frac{\mu^k}{k!}e^{-\mu}.$$

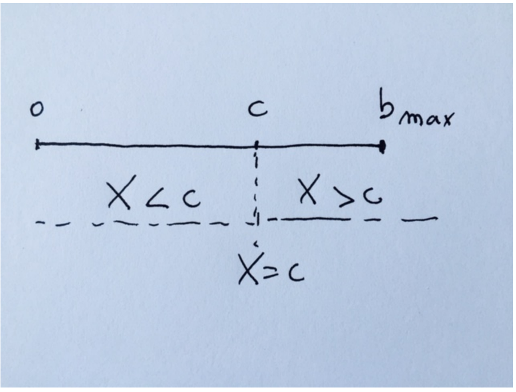

Figure 1: Valori che puó assumere la variabile aleatoria X nel problema del buffer

2. $0 < Y = k < b_{max} - c$:
La probabilitá che al termine di Δt_2 vi siano esattamente $Y = k$ pacchetti con $k = 1{,}2, \cdots, b_{max} - c - 1$ é uguale alla probabilitá che in Δt_1 siano stati bufferizzati $X = k + c$ pacchetti, con $k + c < b_{max}$

$$P_Y(Y = k) = P_X(X = k + c) = \frac{\mu^{k+c}}{(k+c)!} e^{-\mu}$$

3. $Y = b_{max} - c$:
E' il caso in cui in Δt_1 il buffer ŝtato riempito con $X = b_{max}$ pacchetti di cui c sono stati inviati in Δt_2:

$$P_Y(Y = b_{max} - c) = P_X(X = b_{max}) = 1 - \sum_{k=0}^{b_{max}-1} \frac{\mu^k}{k!} e^{-\mu}$$

(b) Qual é la probabilitá che alcuni pacchetti vengano scartati durante il primo slot?

Soluzione: La probabilitá che alcuni pacchetti vengano scartati durante il primo slot é uguale alla probabilitá che piú di b_{max} pacchetti siano stati generati alla sorgente in Δt_1 ovvero

$$P_X(X > b_{max}) = \sum_{k=b_{max}+1}^{\infty} \frac{\mu^k}{k!} e^{-\mu} = 1 - \sum_{k=0}^{b_{max}} \frac{\mu^k}{k!} e^{-\mu}$$

Esercizio 2.

Pippo si reca in banca per effettuare un versamento, e ha uguale probabilitá di trovare 0 o 1 clienti in coda prima di lui. Qualora vi sia effettivamente un cliente, il tempo di servizio é distribuito esponenzialmente con parametro λ.

(a) Qual é la funzione di distribuzione cumulativa (CDF) del tempo di attesa di Pippo?

Soluzione: (Distribuzione esponenziale) Sia X la VA che denota il tempo di attesa e Y il numero di clienti trovati. Chiaramente, per $x < 0$ $F_X(x) = 0$, mentre per $x \geq 0$,

$$F_X(x) = P(X \leq x) = \frac{1}{2} P(X \leq x \mid Y = 0) + \frac{1}{2} P(X \leq x \mid Y = 1). \tag{8}$$

Se non ci sono clienti

$$P(X \leq x \mid Y = 0) = 1,$$

e per specifica di problema, nel caso in cui vi sia una persona

$$P(X \leq x \mid Y = 1) = 1 - e^{-\lambda x},$$

Sostituendo nella (8), la CDF si scrive come

$$F_X(x) = \begin{cases} \frac{1}{2}\left(2 - e^{-\lambda x}\right), & \text{se } x \geq 0, \\ 0 & \text{altrimenti,} \end{cases} \tag{9}$$

ed ha un punto di discontinuitá in $x = 0$.

Esercizio 3.

Siano J e K due variabili aleatorie indipendenti le cui PMF sono definite come segue:

$$P_J(j) = \begin{cases} 0.2 & j = 1, \\ 0.6 & j = 2, \\ 0.2 & j = 3, \\ 0 & \text{altrimenti.} \end{cases} \tag{10}$$

$$P_K(k) = \begin{cases} 0.5 & k = -1, \\ 0.5 & k = 1, \\ 0 & \text{altrimenti.} \end{cases} \tag{11}$$

Si definisca una terza variabile aleatoria M come somma delle due

(a) Trovare il momento ordine tre di M.

Soluzione: (Funzioni generatrici dei momenti)
Per procedere, calcolo le funzioni generatrici dei momenti (FGM) delle due VA:

$$\varphi_J(u) = 0.2e^u + 0.6e^{2u} + 0.2e^{3u}$$

$$\varphi_K(u) = 0.5e^{-u} + 0.5e^u$$

Sia

$$M = J + K$$

la VA di interesse.
La FGM di M é allora:

$$\varphi_M(u) = \varphi_J(u)\varphi_K(u) = (0.2e^u + 0.6e^{2u} + 0.2e^{3u})(0.5e^{-u} + 0.5e^u) = 0.1 + 0.3e^u + 0.2e^{2u} + 0.3e^{3u} + 0.1e^{4u}$$

Per trovare il momento di ordine tre, differenziamo tre volte la FGM di M e ne calcoliamo il valore in $u = 0$:

$$E\left[M^3\right] = \left.\frac{d^3\varphi_M(u)}{du^3}\right|_{u=0} = 16.4$$

(b) Determinare la PMF di M, $P_M(m)$.

Soluzione:
Il valore della PMF $P_M(m)$ in ciascun m é esattamente il coefficiente dell'm-esimo termine e^{mu} in $\varphi_M(u)$:

$$P_M(m) = \begin{cases} 0.1 & m = 0{,}4, \\ 0.3 & m = 1{,}3, \\ 0.2 & m = 2, \\ 0 & \text{altrimenti.} \end{cases} \tag{12}$$