

Tema d'esame di Statistica e analisi dei dati

Prova scritta del 24 gennaio 2018

Esercizio 0

1. Data una popolazione X con valore atteso $\mu = E(X)$ e un campione X_1, \dots, X_n estratto da X , dimostrare che la media campionaria

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

è uno stimatore non deviato di μ .

2. Tracciare il grafico della funzione di distribuzione cumulativa di una variabile aleatoria esponenziale e di una variabile aleatoria geometrica, entrambe di parametro 0.1.
3. Esprimere il parametro λ di una variabile aleatoria esponenziale X in funzione di $E(X)$.
4. Esprimere la deviazione standard di una variabile aleatoria esponenziale X in funzione di $E(X)$.
5. La Figura 1 mostra il grafico della funzione di distribuzione cumulativa di due variabili aleatorie esponenziali: indicare quale dei due grafici corrisponde alla distribuzione con il valore più alto per il parametro.

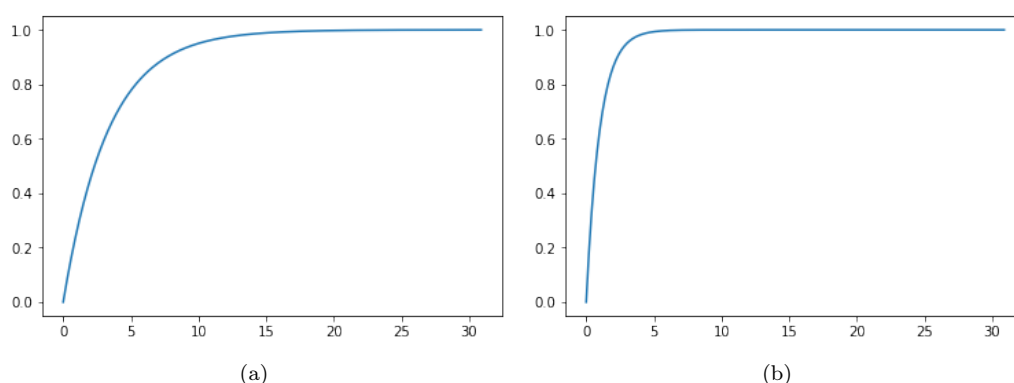


Figura 1: Il grafico della funzione di distribuzione cumulativa di due variabili aleatorie esponenziali.

Esercizio 1

Al porto arrivano barche da pesca che scaricano il loro pesce. Il mercato del pesce ha acquisito una nuova bilancia elettronica che, oltre a pesare il pescato, ha anche un dispositivo ottico che è in grado di identificare la presenza di astici. Ogni volta che il dispositivo segnala la presenza di un astice, registra su un file alcune informazioni, come descritto di seguito. Il file contiene dunque una riga per ogni astice incontrato; inoltre le colonne sono separate dal simbolo "," e i numeri reali sono stati registrati con il simbolo "." come separatore dei decimali.

Per accedere al file prodotto dalla nuova bilancia, collegarsi al sito upload.di.unimi.it, selezionare l'esame di *Statistica e analisi dei dati* e scaricare il file `astici.csv`, che contiene i seguenti attributi:

- *kg.di.pesce*: peso del pescato che contiene l'astice (espresso in kg),
- *settore.di.pesca*: parte di mare in cui è stato pescato l'astice,
- *forza.del.mare*: condizioni del mare durante la pesca,
- *peso.astice*: peso dell'astice trovato (espresso in hg).

1. Quanti astici sono stati trovati?

2. Quanti sono i settori di pesca della parte di mare considerata?
3. Quanti astici sono stati trovati nel settore "A"?
4. Tracciare un grafico opportuno per visualizzare le frequenze dei settori di pesca nel dataset.
5. Qual è stato il settore di pesca nel quale si è trovato il maggior numero di astici?
6. Qual è la percentuale di astici trovati in tale settore?
7. Il carattere *forza.del.mare* è categorico, quantitativo discreto o quantitativo continuo?
8. Tracciare un grafico opportuno per descrivere la *forza.del.mare*.
9. Tracciare il grafico della funzione cumulativa empirica di *forza.del.mare*.
10. La *forza.del.mare* è distribuita in modo normale? Giustificare, possibilmente con un grafico, la risposta.
11. Qual è stata la forza del mare media?
12. Qual è stata la forza del mare riscontrata più frequentemente?

Concentriamoci ora sul peso degli astici trovati.

13. Il carattere *peso.astice* è categorico, quantitativo discreto o quantitativo continuo?
14. Tracciare un grafico opportuno per descrivere il *peso.astice*.
15. Il dataset presenta un outlier per il carattere *peso.astice*: evidenziare tale valore utilizzando un'opportuna rappresentazione grafica.
16. Memorizzare nella variabile `astici.filtrato` il dataset ottenuto eliminando la riga che corrisponde all'outlier identificato al punto precedente.
17. Stimare il valore atteso e la varianza del peso degli astici pescati, escludendo l'outlier.
18. Sempre escludendo l'outlier e sfruttando opportunamente i risultati ottenuti al punto precedente, fornire una maggiorazione della probabilità che il peso di un astice disti dal suo valore atteso, per eccesso o per difetto, almeno 10 hg.
19. Si ritorni ad utilizzare il dataset completo e si controlli graficamente se sussiste una relazione tra:
 - 19.1. *peso.astice* e *forza.del.mare*;
 - 19.2. *peso.astice* e *kg.di.pesce*.
20. Nel caso in cui si ritiene che ci sia una dipendenza, precisare di che tipo di dipendenza si tratta e calcolare un indice numerico che stabilisca l'intensità di tale relazione.
21. I pescatori vorrebbero qualche garanzia riguardo alla stima del valore atteso del *peso.astice*. In particolare, noto il numero di astici pescati (è stato richiesto al punto 1. di questo esercizio), vorrebbero che la probabilità di errore nella stima fosse almeno uguale a 0.9. Qual è allora l'errore (espresso in hg) che devono essere disposti a tollerare?

Esercizio 2

Concentriamoci ora sul carattere *kg.di.pesce*.

1. Tracciare la funzione cumulativa empirica del carattere *kg.di.pesce*.
2. Stimare il valore atteso e la varianza dei *kg.di.pesce*.
3. Gli stimatori utilizzati al punto precedente sono non devianti? Giustificare la risposta.
4. C'è un modello probabilistico noto che ben descrive il carattere *kg.di.pesce*: quale potrebbe essere? Giustificare la risposta.
5. Stimare il parametro che caratterizza il modello identificato al punto precedente.