

SAD Esame Giugno 2023

ESERCIZIO 1

Sistema di home banking che ha password formate da 5 cifre decimali, si ipotizzi che la nostra password sia p .

1. Quante sono le password possibili?
2. Qual è la probabilità che una password estratta a caso sia uguale a p ?
3. Qual è il numero di password distinte che contengono la stessa cifra nelle prime 3 posizioni?
4. Qual è la Probabilità di 1.2 condizionata da 1.3 (scegliendone una a caso tra tutte quelle che iniziano con tre cifre uguali)
5. Numero di password distinte che hanno 3 cifre consecutive uguali in qualsiasi posizione (non solo all'inizio)
6. Calcola la probabilità di 1.4, senza il vincolo che le 3 cifre uguali e consecutive siano all'inizio, ma ovunque

ESERCIZIO 2

$X \sim G(h)$ variabile aleatoria con distribuzione geometrica con parametro h

1. Valori di x per cui ha senso calcolare $P(X > x)$
2. Calcola $P(X > x)$ in funzione di h
3. Scrivi $F_X(x)$ (funzione di ripartizione) in funzione di x e h
4. Grafico di F_X con $h = 4/10$

ESERCIZIO 3

Con $n \in \mathbb{N}$, siano X_1, X_2, \dots, X_n campione di variabili aleatorie indipendenti e identicamente distribuite secondo modello geometrico di parametro h . La distribuzione della popolazione è la stessa di X dell'esercizio 2. Si consideri $\mu = E(X)$ e $T = (1/n) * \sum_{i=1}^n X_i$

1. T è uno stimatore non distorto per μ ? Motiva
2. T è uno stimatore non distorto per h ? Motiva
3. Calcola bias e scarto quadratico medio di T con stima del parametro h ?
4. Lo stimatore T per il parametro h , gode della proprietà di consistenza in media quadratica?
5. Fissato $\varepsilon > 0$, esprimere in funzione di ε , h e n la probabilità dell'evento che si verifica quando l'errore (in valore assoluto) che si compie usando T per stimare μ sia migliore o uguale a ε , giustificando i passaggi e individuando eventuali approssimazioni necessarie
6. Proporre uno stimatore S per h , giustificando la scelta fatta

ESERCIZIO 4

Usando il csv `banking.csv` che contiene le colonne `utenti`, `minuti`, `illeciti` e `tentativi` (?)

1. Quanti utenti distinti ci sono nel dataset e quali sono gli attributi per i valori mancanti?
2. Valutare l'ipotesi che i valori di *minuti* possano essere assimilati in un campione estratto da una popolazione distribuita secondo modello normale
3. Visualizza grafico per *minuti* e motiva la scelta, utilizzando anche la risposta alla domanda 4.2
4. Esiste una relazione tra *minuti* e *illeciti*? Se sì quale e con che strumenti la vedo?
5. Descrivi l'attributo *tentativi* con il grafico che ritieni migliore e motiva la scelta
6. Memorizza in una variabile *dati_illeciti* tutti i casi illeciti e visualizza il graficamente i corrispondenti valori dell'attributo *tentativi*
7. Sulla base di 4.6, considera ogni modello di distribuzione specificando se e perché, si tratta di modelli ragionevoli per i valori in *dati_illeciti*
 - a. Modello esponenziale
 - b. Modello gaussiano
 - c. Modello geometrico
 - d. Modello uniforme discreto

ESERCIZIO 5

Lavoriamo sui tentativi relativi ad accessi illeciti e saranno i valori assunti da un campione estratto da una popolazione distribuita come una variabile aleatoria Y :

1. Ci sono valori mancanti per l'attributo? Se sì, quanti?
2. Stima di $E(Y)$, specificando la dimensione del campione utilizzato e indicando varie proprietà dello stimatore utilizzato
3. I dati a disposizione permettono di garantire, con probabilità maggiore o uguale a 0.99, che la stima fatta in 5.2 comporti un errore (in valore assoluto) di massimo $\varepsilon = 0.1$
4. Ipotizziamo $Y \sim G(h)$ e che h è sconosciuto, proponi uno stimatore
5. Lo stimatore di 5.4 gode di qualche proprietà?
6. Applica stimatore di 5.4 per ottenere una stima di \hat{h}

Calcola le seguenti probabilità facendo riferimento al modello assunto per Y e sostituendo a h la stima \hat{h} di 5.6

7. Calcola $P(3 < X < 8)$, che è la probabilità che il numero di tentativi di autenticazione durante un accesso illecito sia strettamente compreso tra 3 e 8
8. Un malintenzionato ha effettuato due tentativi di accesso a vuoto. Calcola la probabilità che siano necessari più di due ulteriori tentativi