

# Orali del 29 settembre 2020

Alessandro Di Gioacchino

5 febbraio 2021

Stimatore non distorto. Che differenza c'è tra scrivere  $X$  ed  $x$ ? Stiamo calcolando il valore atteso di cosa? Qual è l'argomento dello stimatore? Dal punto di vista pratico, cosa vuol dire che lo stimatore è non deviato? Se uso uno stimatore non deviato, c'è errore nella stima? Esempio dei due campioni (uguale a quello riportato nel file dell'orale del 2020.09.30).

Cos'è il valore atteso? Perché è interessante il concetto di valore atteso? Cosa mi racconta? Che informazioni mi dà? Il valore atteso coincide sempre con una specificazione della variabile aleatoria? Facciamo un esempio in cui non accade.

Quale distribuzione possiamo usare per modellare il lancio di un dado? Uniforme discreta. Grafico della funzione di ripartizione. Fino a che punto sale il grafico a scalini? Evidenziamo nel grafico  $P(X < \frac{7}{2})$

Quale conclusione possiamo quindi trarre? A quale concetto possiamo collegarci? (Non so che intendesse qui)

Quale significato ha il fatto che lo stimatore di una quantità ignota sia non distorto? Perché ci piace che uno stimatore sia non distorto quando lo usiamo per stimare qualcosa che non conosciamo? Il fatto che lo MSE sia uguale alla varianza è poco interessante in questo ambito; è molto interessante se stessi parlando di consistenza. Anche se il valore atteso dello stimatore è uguale alla quantità sconosciuta, non abbiamo alcuna garanzia che la stima ottenuta a partire dallo stimatore sia vicina alla quantità da stimare.

Valore atteso di una variabile aleatoria discreta. Quali  $x$  sono coinvolte dalla definizione? Cosa si intende per 'supporto'? Le  $x$  sono tutte le specificazioni della variabile aleatoria; che significato ha la funzione  $p_X$ ? Il valore atteso è un numero: che informazione mi fornisce? Dà un'informazione sulla centralità della distribuzione. Cosa significa, per esempio, che 42 è un valore centrale per la distribuzione di una variabile aleatoria? Tendo a vedere i miei dati in un intorno di 42

Un intorno piccolo o grande? Dipende dalla varianza. Come calcoliamo il valore atteso di una funzione in più variabili aleatorie? Cosa cambierebbe nel significato del numero ottenuto? Cosa ci sarebbe nell'intorno di 42 considerato sopra?

Proprietà formali del valore atteso.

Eterogeneità. Quali casi di massima eterogeneità ci sono? Per quale tipo di dati ha senso parlare di omogeneità ed eterogeneità? Per dati qualitativi: non ha senso calcolare l'omogeneità associata all'altezza delle persone. Abbiamo un campione di 5 elementi: rosso, verde, rosso, rosso, blu.  $n = 5$ : quanto vale la frequenza di 'rosso'? Quanto la frequenza di 'verde'? Quanto la frequenza di 'rosso'? (Di nuovo) L'indice di Gini considera il numero di valori osservabili, e in particolare le loro frequenze relative. Perché cattura il concetto di eterogeneità? Se tutte le frequenze relative sono uguali, non è detto che tutti gli elementi del campione hanno valori diversi: cosa può capitare? Ad esempio, se ho 3 etichette con frequenza relativa  $\frac{1}{3}$ , sono tutte equamente rappresentate. Calcolo l'indice di Gini associato ad un campione ed ottengo 0.44: quali conclusioni posso trarre? Quali conclusioni posso trarre se ottengo 0.1? Dipende dal numero di osservazioni, motivo per cui di solito si usa l'indice di Gini normalizzato.

Stimatori non distorti. Cosa si intende quando diciamo che  $\Theta$  è un parametro? C'è differenza tra avere uno stimatore non distorto oppure un bias pari a zero?

Perché è desiderabile che uno stimatore sia non distorto? Il fatto che il valore atteso dello stimatore sia uguale a quanto voglio stimare implica che la stima sia precisa? Cosa si intende per 'precisa'? Uno stimatore corretto non mi garantisce che il valore della stima sia vicino o lontano al valore che voglio stimare: se misurassimo la vicinanza con il quadrato della differenza, uno stimatore deviato potrebbe paradossalmente restituire qualcosa di più vicino alla quantità ignota rispetto ad uno stimatore non deviato.

Quale concetto di statistica inferenziale abbiamo visto, che è più legato alla vicinanza o lontananza dalla quantità stimata? Di cosa parla la consistenza? Riguarda il limite per che cosa? Cosa vuol dire calcolare la varianza di qualcosa che tende ad infinito? (Domanda retorica, il candidato stava sbagliando la definizione di MSE)