

Orali del 17 giugno 2020

Alessandro Di Gioacchino

5 febbraio 2021

Abbiamo un campione composto da coppie: come costruiamo un diagramma di dispersione? Cosa indica un generico punto del grafico? A cosa serve un diagramma di dispersione? Quali indici quantitativi abbiamo visto che ci permettono di confermare o smentire un'ipotesi avanzata dopo aver generato il diagramma?

Cos'è l'indice di correlazione? Cosa sono X ed Y ? Due campioni, non due variabili aleatorie. Perché è rilevante conoscere il codominio dell'indice di correlazione? Cosa posso dire sulla relazione se l'indice di correlazione vale 1? È lineare diretta. Cosa posso dire se calcolassi il coefficiente di correlazione lineare (cioè l'indice di correlazione) ed ottenessi un valore molto vicino a zero?

Com'è definita la covarianza? Perché questo indice è meno 'lampante'? Dipende molto dall'unità di misura.

Com'è definita la covarianza tra due variabili aleatorie? Dimostrazione (pagina 133):

$$\mathcal{E}[(X - \mu_X)(Y - \mu_Y)] = \mathcal{E}[XY] - \mathcal{E}[X]\mathcal{E}[Y]$$

Anche se il valore atteso è un operatore lineare, non è vero che $\mathcal{E}[(X - \mu_X)(Y - \mu_Y)] = \mathcal{E}[X - \mu_X]\mathcal{E}[Y - \mu_Y]$

Il valore atteso di qualcosa è una costante, ed il valore atteso di una costante è la costante stessa.

Quando due variabili aleatorie si dicono 'indipendenti'? Come scegliamo gli insiemi di numeri reali A e B ? (Pagina 113)

L'equazione deve valere per qualunque coppia di insiemi, non necessariamente disgiunti. Le due variabili aleatorie non devono neanche indicare la stessa cosa. Che relazione legata al valore atteso si può ricavare a partire da questa definizione? Cosa ci permette di dire sulla covarianza il fatto che il valore atteso del prodotto delle v.a. è uguale al prodotto dei valori attesi? Cosa implica cosa? Partiamo dall'ipotesi che X ed Y siano indipendenti: cosa possiamo quindi dedurre? Cosa possiamo dedurre dall'equazione $\mathcal{E}[XY] = \mathcal{E}[X]\mathcal{E}[Y]$? Quindi indipendenza implica nullità della covarianza. Posso affermare anche il viceversa? No, esistono alcuni controesempi.

Tornando all'indice di correlazione fra due campioni, cosa posso azzardare qualora questo dovesse essere vicino a zero? Non possiamo essere sicuri che i due campioni siano indipendenti, ma il dubbio dovrebbe venirci.

Modello geometrico. Conta il numero di esperimenti bernoulliani prima di un successo a caso, o prima del primo successo? Come sono gli esperimenti bernoulliani? Come ricaviamo la funzione di massa di probabilità nella sua forma analitica? Quanti fallimenti conto prima del primo successo? x tentativi con un insuccesso, un tentativo con un successo. Aggiungiamo anche la funzione indicatrice. Qual è l'insieme alla base di questa funzione indicatrice? Perché il parametro p del modello non può essere zero? Ci vorrebbero infiniti esperimenti. Grafico della funzione di ripartizione, con $p = 0.1$

In particolare, come approssima il valore 1 sull'asse delle ordinate? La lunghezza dei segmenti paralleli all'asse delle ascisse è variabile o è sempre la stessa? È sempre uguale ad 1

In quale intervallo posso scegliere il valore di p ? Tra zero (escluso) ed uno (incluso). Cosa succede facendo scendere p verso zero pian piano? Come sarebbe lo stesso grafico per p molto vicino ad 1?

Quale macro-argomento non abbiamo ancora affrontato? La statistica inferenziale. Ho un campione X_1, \dots, X_n estratto da una popolazione X con valore atteso $\mathcal{E}[X] = \mu$

Sono interessato a stimare la varianza della popolazione con lo stimatore $T = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$

Esiste un unico stimatore non distorto per un dato parametro? Come verifichiamo se lo stimatore è distorto o meno? Calcoliamo il valore atteso di T

Di cosa stiamo calcolando il valore atteso in $\mathcal{E}[(X_i - \mu)^2]$? Come si definisce un campione casuale? Cos'è μ ? Essendo il campione distribuito come la popolazione, μ è anche il valore atteso del campione. Stiamo quindi calcolando il valore atteso di una variabile aleatoria meno il suo valore atteso, al quadrato: si tratta della varianza di X_i

Come eliminiamo la dipendenza da i in $\frac{1}{n} \sum_{i=1}^n \mathcal{V}[X_i]$? Quindi lo stimatore non è distorto. Perché all'inizio avevamo pensato che lo fosse? Che altra differenza c'è tra T e la varianza campionaria? Scrivendo uno stimatore, possiamo ragionarci sia pensando che le X_i siano v.a., sia pensando che siano valori. La prima è che c'è $n-1$ al denominatore, poi ce n'è un'altra che compensa tale differenza. Qual è la definizione formale di stimatore? È una funzione di un

campione e di cos'altro? Solo di un campione, non può dipendere da parametri incogniti. Perché non avrebbe senso altrimenti? La stima che otterrei sarebbe comunque sconosciuta. La "varianza campionaria" scritta così:

$$S = \frac{1}{n-1} \sum_{i=1}^n (X_i - \mathcal{E}[X])^2$$

dipende solo dal campione? No, anche dal valore atteso della popolazione. La varianza campionaria non dipende dal valore atteso della popolazione, ma dalla media campionaria.