

# Orali del 8 settembre 2020

Alessandro Di Gioacchino

5 febbraio 2021

Abbiamo un campione di valori: come posso trasformarlo in modo che il massimo sia uguale ad 1? Perché scrivere  $y := \frac{x}{h}$  non è proprio corretto? Perché  $x$  è un insieme.

Ragioniamo ora in termini di variabili aleatorie:  $X$  descrive la popolazione da cui il campione precedente è stato estratto. Conoscendo la distribuzione di  $X$ , cosa posso dire qualcosa sulla distribuzione di  $Y$ ? Supponiamo che  $X \sim B(p)$ , ed  $Y$  è la sua standardizzazione.  $Y = \frac{X-\mu}{\sigma}$ : cosa sono  $\mu$  e  $\sigma$  nel caso specifico della Bernoulli? Che forma assume  $Y$ ? Se  $X$  ed  $Y$  avessero la stessa distribuzione, anche  $Y$  dovrebbe essere una Bernoulli: cosa caratterizza una distribuzione del genere? Le specificazioni di una Bernoulli sono 0, 1

Quali sono le specificazioni di  $Y$ ? Quindi può essere anche lei di Bernoulli? Perciò non è vero che, sommando o moltiplicando una variabile aleatoria per una costante, la distribuzione originale viene mantenuta dalla nuova variabile. Ci sono casi particolari in cui questo vale, o comunque in cui riesco a dire qualcosa della distribuzione associata alla nuova variabile? Riflettiamo sulla funzione di ripartizione di  $Y$  (in generale, non è più una Bernoulli): cosa c'è sull'asse delle ordinate? Cosa c'è sull'asse delle ascisse? Possiamo esprimere la funzione di ripartizione di  $Y$  in funzione di quella di  $X$ , con  $Y = \frac{X}{h}$ ? Possiamo aspettarci che  $X$  ed  $Y$  abbiano lo stesso valore atteso? No, essendo il valore atteso un operatore lineare. Quanto vale  $\lim_{x \rightarrow \infty} F(x)$ ? 1

È possibile che l'asintoto orizzontale diventi la retta  $x = \frac{1}{h}$ ? No.

Come costruiamo un diagramma di dispersione? Dobbiamo necessariamente calcolare l'indice di correlazione? No. Come sempre si parte da un campione: come è fatto? Come genero il grafico a partire dai suoi valori?

Come disegniamo un diagramma di dispersione a mano? Generando questo grafico, a cosa sono interessato? Dobbiamo scegliere quale elemento delle coppie inserire sull'asse delle ascisse e quale sull'asse delle ordinate: se invertissimo questa scelta, il comportamento del diagramma di dispersione sarebbe lo stesso o no? È ovvio che il diagramma cambia, ma la sua interpretazione no. L'assenza di relazioni tra variabili aleatorie è stata formalizzata attraverso un concetto importante: quale? L'indipendenza. Quando diciamo che due variabili aleatorie sono indipendenti? Grazie a questa definizione possiamo dire qualcosa sulla funzione di massa di probabilità di variabili aleatorie: cosa?  $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ : come si chiamano quelle al secondo membro? Funzioni di massa marginali. Essendo la massa di probabilità congiunta pari al prodotto delle marginali per variabili aleatorie indipendenti, possiamo dimostrare alcune cose legate al valore atteso. Come dimostriamo che il valore atteso di due v.a. indipendenti è pari al prodotto dei singoli valori attesi? Qual è la definizione di valore atteso calcolato su una funzione di due variabili aleatorie? Cosa mi permette di dire la linearità del valore atteso? Possiamo estendere questa proprietà a qualunque funzione?

Abbiamo una popolazione  $X$  ed un campione  $X_1, \dots, X_n$ : vogliamo stimare il valore atteso di  $X$

Come è fatto lo stimatore "media campionaria"? Perché gode di correttezza e consistenza? Perché queste proprietà sono desiderabili per uno stimatore? Nel nostro caso quale parametro vogliamo stimare? Cos'è  $\tau$ ? Cos'è  $\Theta$ ? Il valore atteso di una popolazione può essere o meno un parametro: dipende dalla distribuzione. Per il momento indichiamo con  $\mu$  il valore atteso di  $X$

Perché è bello che uno stimatore sia non distorto rispetto a quanto sta stimando? Il fatto che il valore atteso dello stimatore è uguale a quanto vogliamo stimare indica che il risultato ottenuto calcolando lo stimatore è vicino alla quantità ignota? La proprietà di non oscillare troppo attorno alla quantità ignota è implicata dal fatto che lo stimatore è non deviato? No, ma dalla sua eventuale consistenza. Cosa mi dice il fatto che lo stimatore sia non deviato? Come dimostriamo che il valore atteso della media campionaria è uguale al valore atteso della popolazione? Perché il valore atteso di  $X_i$  è uguale al valore atteso di  $X$ ? Essendo identicamente distribuite, hanno tutte lo stesso valore atteso: qual è il valore atteso uguale per tutte le  $X_i$ ? Perché? Come sono distribuite le  $X_i$ ? Il teorema centrale del limite si applica alla somma delle  $X_i$ , mentre noi siamo interessati al valore atteso della singola  $X_i$

Le variabili aleatorie che compongono un campione casuale devono essere indipendenti, identicamente distribuite, e seguire la stessa distribuzione della popolazione: altrimenti non potrei dire di aver estratto il campione da quella

popolazione. Quindi il valore atteso delle  $X_i$  è uguale al valore atteso di  $X$

A cosa serve un box plot? Come definiamo la mediana o, più in generale, i quartili? Un box plot serve a farmi un'idea sulla distribuzione dei dati. Che idea potrei farmi conoscendo di un campione solo il suo box plot? Se il box plot fosse asimmetrico a sinistra, da quale modello potrebbe essere stato estratto il campione che ha generato tale box plot?

Distribuzione esponenziale. Immaginiamo che  $X \sim E(\lambda)$ , e definiamo  $Y := \alpha X$ : possiamo dire qualcosa sulla distribuzione di  $Y$ ? Se  $\alpha$  fosse minore di 0, il dominio di  $Y$  cambierebbe e di conseguenza anche la distribuzione che segue. Se vincolassi  $\alpha > 0$ ? Ci conviene ragionare in termini della funzione di ripartizione: calcoliamo quella di  $Y$ . Come definiamo la funzione di ripartizione di  $Y$ ? A quale conclusione possiamo arrivare partendo da questa definizione? La funzione di ripartizione non resta la stessa: sostituiamo ad  $Y$  la sua definizione, cioè  $\alpha X$ , ma l'argomento della funzione di ripartizione  $y$  resta uguale. Come possiamo manipolare questa espressione per ricondurci a qualcosa che sappiamo calcolare? Sappiamo che  $X$  ha una distribuzione esponenziale, quindi dovremmo ricondurci alla sola  $X$  per capire di quale evento stiamo calcolando la probabilità:  $P(\alpha X \leq y) = P(X \leq \frac{y}{\alpha})$ , perché  $\alpha > 0$

A che probabilità ci siamo ricondotti?  $P(X \leq \frac{y}{\alpha}) = 1 - e^{-\lambda \frac{y}{\alpha}}$

Come trasformiamo questa roba per ottenere la forma analitica di una funzione di ripartizione nota? Una funzione di ripartizione individua in modo univoco la distribuzione, quindi se riconoscessi in quanto scritto sopra la funzione di ripartizione di un particolare modello potrei dire che  $Y$  segue tale modello. Quella sopra è la funzione di ripartizione di un'esponenziale con parametro  $\frac{\lambda}{\alpha}$

Quindi resto nel modello esponenziale pur avendo applicato una scalatura ad  $X$

Perché avere uno stimatore non distorto per una quantità ignota è una proprietà desiderabile? Il fatto che il valore atteso dello stimatore è uguale a quanto voglio stimare non descrive l'ampiezza delle oscillazioni della mia stima. Cosa mi implica dire che il valore atteso dello stimatore è quello che voglio stimare? Perché è bello che uno stimatore sia non distorto? Perché quanto otteniamo è una buona stima? Cosa significa che il valore atteso di una variabile aleatoria è 42?

Abbiamo a disposizione un campione di dati, e vogliamo valutare l'ipotesi che sia stato estratto da una distribuzione normale: come possiamo farlo? Una possibilità è disegnare l'istogramma dei dati, ma c'è un grafico specifico per effettuare questo test. Avendo un campione di  $n$  elementi, come tracciamo un diagramma QQ? Quali quantili della distribuzione andiamo a calcolare? Gli stessi calcolati a partire dal campione. Perché è ragionevole dire che, se i punti tendono ad allinearsi sulla retta  $y = x$ , il campione è estratto da una normale? Perché i quantili campionari approssimano quelli teorici, ma anche perché conoscendo i quantili di una distribuzione so tutto della distribuzione: l'insieme di tutti i quantili individua in modo univoco la distribuzione. Per calcolare i quantili teorici, dobbiamo istanziare la normale con alcuni parametri: da dove li andiamo a prendere? Potrei stimarli, o normalizzare la normale in modo da usare direttamente i quantili della normale standard. Come stimiamo  $\mu$  e  $\sigma$ ? Perché scegliamo la media e la varianza campionarie? Perché è bello che l'approssimazione non abbia un bias? Nella definizione di stimatore non distorto non prendiamo mai in considerazione la taglia del campione. Non è detto che la stima fatta con uno stimatore non deviato sia particolarmente buona. Cosa vuol dire che il valore atteso di una variabile aleatoria  $X$  è 42? Il valore atteso di una Bernoulli è un numero tra 0 ed 1, per cui non posso dire che  $X$  debba assumere il suo valore atteso. Il valore atteso è un indice di che cosa? Di centralità: quindi? Le specificazioni della v.a. si trovano attorno al valore atteso, più o meno vicini ad esso: dire che lo stimatore è non deviato implica che i valori assunti dallo stimatore girano attorno al valore che voglio stimare.

Immaginiamo di avere un campione a coppie, e di aver calcolato le frequenze congiunte; ciascun elemento della coppia può valere 0 oppure 1

Costruiamo una tabella delle frequenze congiunte (con numeri a caso); nelle righe abbiamo i valori delle  $x$ , mentre nelle colonne quelli delle  $y$ : come calcoliamo le frequenze marginali delle  $x$ ? Cosa deve contenere la frequenza marginale di  $x$  per il valore 0? Il numero di osservazioni in cui  $x$  ha preso il valore 0

Qual è la frequenza marginale di  $x$  rispetto al valore 1? Cosa cambia per calcolare le frequenze marginali di  $y$ ? Ragionando in termini della tabella, sarebbe bastato sommare sulle righe nel primo caso e sommare sulle colonne nel secondo.

Abbiamo due variabili aleatorie  $X$  ed  $Y$ , di cui conosciamo la funzione di massa congiunta  $p_{X,Y}$ : come otteniamo la funzione di massa  $p_X$ ? Come calcolo  $p_X(x_i)$  a partire da  $p_{X,Y}$ ? Per calcolare  $p_X(6)$ , questo 6 rimane fissato in  $p_{X,Y}$  mentre l'altro argomento no: poi come procedo? Come calcolo  $p_X(6)$  mettendo solamente in campo la funzione di massa congiunta  $p_{X,Y}$ ? Come abbiamo calcolato prima le frequenze marginali del campione? Cosa abbiamo fatto dopo aver calcolato le frequenze (non ancora marginali)? Le abbiamo sommate. Quindi  $p_X(6) = \sum_y p_{X,Y}(6, y)$ : basta sommare per la congiunta al variare di tutti i possibili valori delle  $Y$

Cos'era un generico elemento della precedente tabella? Una frequenza congiunta.

Cambia qualcosa di rilevante se le v.a. fossero state continue? No, sarebbe bastato valutare l'integrale.