

Orali del 30 settembre 2020

Alessandro Di Gioacchino

5 febbraio 2021

(Il primo orale è parziale)

Cosa descrive, in generale, una distribuzione di cui non conosciamo un parametro? Il nostro campione e la nostra popolazione: il campione segue la stessa distribuzione della popolazione, altrimenti non starei campionando bene.

Cos'è $\tau(\Theta)$? È una quantità ignota, non necessariamente uguale a Θ

Esempio in cui siamo interessati a $\tau(\Theta) \neq \Theta$

Il valore atteso di un numero è diverso dal numero stesso? (Domanda retorica, la risposta è no) Cosa stimiamo tipicamente? Il valore atteso di una popolazione.

Volendo stimare il valore atteso μ di una popolazione normale, cosa sono Θ e $\tau(\Theta)$? Entrambi μ

A cosa è uguale \bar{X} ? Perché posso portare fuori dal valore atteso una costante moltiplicativa? Per la sua linearità. Perché il valore atteso delle X_i , estratte dalla popolazione X , è uguale al valore atteso di X ? Ci interessa che le X_i siano indipendenti o identicamente distribuite? (Ci interessa parzialmente il fatto che siano identicamente distribuite, ma l'aspetto importante è che seguono la stessa distribuzione della popolazione) Variabili aleatorie indipendenti hanno lo stesso valore atteso? Come sono distribuite le X_i ? Come la popolazione. Perché è bello che uno stimatore sia non deviato, cioè che il suo valore atteso sia uguale a $\tau(\Theta)$? Se uno stimatore non distorto mi desse esattamente la quantità ignota, come tengo conto del fatto che due campionamenti distinti a cui applico lo stimatore danno risultati diversi? (Esempio dei due campioni menzionato nel file dell'orale del 29.09.2020)

Il valore atteso della popolazione è legato a quello del campione? Ogni elemento ha valore atteso μ , ma non è vero che ogni elemento è uguale a μ

Funzione di ripartizione di una geometrica. La 'e' sull'asse delle ordinate è maiuscola o minuscola? Perché la x sull'asse delle ascisse è minuscola? Perché sono le specificazioni della variabile aleatoria. Quali specificazioni ha una geometrica? \mathbb{N}^+

Quanto vale la funzione di ripartizione in 0? A che evento corrisponde? Posso dover fare 0 tentativi prima di ottenere il primo successo? (Dipende se stiamo includendo o no il primo successo nella conta degli esperimenti: se sì, è ovvio che per ottenere un successo devo quantomeno fare un esperimento; se viceversa contassimo solo il numero di fallimenti che precedono il primo successo, allora c'è una certa probabilità p che il primo esperimento sia un successo, quindi la variabile aleatoria assume il valore 0) Qual è l'esperimento alla base della geometrica? Includiamo o no il primo successo tra gli insuccessi?

Abbiamo un campione di coppie: $(1, 0), (1, 3), (3, 1), (1, 0), (3, 1)$

Calcoliamo le frequenze congiunte di questo campione, e visualizziamole in un modo opportuno.

Concetto di probabilità condizionata. La definizione vale sempre? Cosa misura questo rapporto?

Ha senso parlare della "frequenza condizionata" di, per esempio, $y \mid x$? (Riferendosi al campione di cui sopra)

Sì: immaginiamo di avere un campione di coppie, in cui un elemento codifica il genere di una persona e l'altro il suo stipendio annuale; in questo caso, condizionare lo stipendio rispetto al genere ha senso? Potrei vedere che mediamente le donne hanno uno stipendio minore degli uomini. Calcoliamo la frequenza condizionata di $y \mid x$ nel campione di cui sopra: possiamo ricavarla dalla tabella delle frequenze congiunte? Quali teoremi abbiamo visto circa la probabilità condizionata? Di Bayes e delle probabilità totali. Dimostriamo il teorema delle probabilità totali dopo averlo enunciato. Quali sono le ipotesi? Sviluppando il calcolo, ottengo varie probabilità condizionate: devo stare attento che queste siano calcolabili (in altri termini, l'evento condizionante non può essere l'evento certo o l'evento impossibile).

Voglio stimare la varianza di una popolazione. Cosa indica $\tau(\Theta)$? Cosa indica Θ ?

Voglio stimare il valore atteso di una popolazione distribuita secondo un'esponenziale di parametro λ : in questo caso specifico, cos'è Θ ? $\Theta = \lambda$

Cos'è $\tau(\Theta)$? $\frac{1}{\lambda}$, il valore atteso dell'esponenziale: volendo stimarlo, cosa mi aspetto dallo stimatore? Se è non deviato, cosa succede? A cosa è uguale il valore atteso dello stimatore? Quanto vale il valore atteso di un'esponenziale?

Quindi, cos'è τ ? (Una funzione che associa $\frac{1}{\Theta}$ a Θ)

A cosa serve un diagramma QQ? Cosa si intende per 'quantile'? Cosa si intende per "quantile teorico"?

Conoscere tutti i quantile di una distribuzione vuol dire averla fissata.

Come si definisce un "quantile teorico"? Come possiamo calcolarli? Così come per ogni $0 \leq q \leq 1$ possiamo definire il q -esimo quantile campionario, per ogni $0 \leq q \leq 1$ possiamo definire il q -esimo quantile di una data distribuzione. Che forma assume la funzione di ripartizione associata ad un'esponenziale con parametro λ ? Che succede al variare di λ ? Il parametro di un'esponenziale può essere un valore qualsiasi? No, deve essere maggiore di zero: quindi, al crescere di λ , $e^{-\lambda x}$ diventa più piccolo mentre $1 - e^{-\lambda x}$ diventa più grande. Individuiamo un valore y compreso tra 0 ed 1 sull'asse delle ordinate, e cerchiamo un valore x tale che y è il valore ottenuto calcolando la funzione di ripartizione con x come argomento; in altri termini stiamo invertendo la funzione di ripartizione. Supponiamo che $y = \frac{1}{2}$: applicando l'inversa della funzione di ripartizione, che numero otteniamo? Calcolando la funzione di ripartizione in un certo punto, di quale evento voglio ottenere la probabilità? Cosa succede mettendo al posto di x (in $P(X \leq x)$) il valore ottenuto in precedenza? (In pratica abbiamo calcolato il secondo quartile dell'esponenziale, cioè la sua mediana) Cambia qualcosa se la funzione di ripartizione calcolasse $P(X \leq x)$ piuttosto che $P(X < x)$? No, perché l'esponenziale è una distribuzione continua. Abbiamo quindi dimostrato che $P(X \leq \frac{\ln 2}{\lambda}) = \frac{1}{2}$: questa osservazione è collegata ad un interessante indice di centralità, la mediana.

Vogliamo stimare la varianza di una popolazione. Che stimatore useremmo? Cosa si intende per "buon stimatore"? Cosa indica la proprietà di correttezza associata ad uno stimatore? Nel nostro caso, quanto vale $\tau(\Theta)$? (Sicuramente $\tau(\Theta)$ è la varianza della popolazione. Non possiamo dire nulla su Θ finché non sappiamo quale distribuzione modella al meglio la popolazione) Cosa rappresenta la variabile aleatoria X ? La popolazione. La lettera usata per indicare uno stimatore è maiuscola o minuscola? Perché siamo interessati agli stimatori non distorti?

Disposizioni. Ho n oggetti e k posti: voglio trovare tutti i modi possibili di disporre gli oggetti nei posti, senza ripeterli più di una volta. Perché $n!$ indica il numero di possibili permutazioni di n oggetti? Lo stesso discorso si applica alle disposizioni: posso occupare il primo posto in n modi, il secondo in $n - 1$, eccetera; arrivato al k -esimo posto, ho $n - k + 1$ scelte

Disposizioni con ripetizione.

Quale semplice domanda di probabilità coinvolge le disposizioni (semplici o con ripetizione)? Calcoliamo ad esempio la probabilità di ottenere 6 lanciando due dadi e sommandone il risultato. Come colleghiamo il numero di casi possibili al concetto di disposizione? Sono 6 oggetti su 2 posti, quindi al denominatore (casi possibili) abbiamo $\binom{6}{2}$. Se al numeratore inserissi le disposizioni senza ripetizione di 6 oggetti in 2 posti, starei calcolando la probabilità di quale evento? Che l'esito dei due lanci sia diverso.

Vogliamo stimare la varianza di una popolazione. A cosa è uguale $\tau(\Theta)$? Cosa sono Θ e $\tau(\Theta)$? Come si definisce formalmente uno stimatore? Normalmente questa definizione pone l'enfasi sul fatto che lo stimatore deve essere funzione del campione e nient'altro.

Usiamo il seguente stimatore:

$$T^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

Dove μ è il valore atteso della popolazione X

Cosa possiamo dire a pelle di questo stimatore? Perché ricorda la varianza campionaria? Ha senso usare T^2 al posto della varianza campionaria per stimare la varianza di una popolazione? Calcoliamo $\mathcal{E}(T^2)$

Ricorda che μ è il valore atteso di X , ma le X_i sono estratte da X : quindi seguono la stessa distribuzione della popolazione, ed hanno il suo stesso valore atteso. Definizione di campione aleatorio. Come sono le variabili aleatorie che compongono il campione? Sono identicamente distribuite (ed indipendenti). Quanto vale la varianza di X_i ? X_i ha la stessa varianza della popolazione, che abbiamo indicato con σ^2

La varianza di X_i è indipendente da i , ed è sempre uguale ad σ^2 : sommando questo valore n volte, ottengo $n\sigma^2$

Immaginiamo di giocare alla roulette, supponendo che possa uscire un numero da 0 a 36; decido inoltre di puntare sullo 0: che modello possiamo utilizzare per determinare se vinco o perdo? Usiamo un modello di Bernoulli, di parametro $p = \frac{1}{37}$

Cambierebbe qualcosa se puntassi su un altro numero? No. Supponiamo di voler giocare finché non vinco: quale modello possiamo utilizzare per questa situazione? Geometrico. Quali sono gli aspetti più importanti di questo modello? Scriviamo la forma analitica della funzione di massa. Calcolando $f(3)$, sto ragionando in termini di quale possibile esito dell'esperimento?

Il modello geometrico si può descrivere in due modi differenti, quindi $f(3)$ può rappresentare due fallimenti seguiti da un successo oppure tre fallimenti prima del primo successo.

Perché $f(x) = p(1-p)^{x-1}$? (Questa è la funzione di massa associata ad un modello geometrico che include il primo successo nella conta degli esperimenti) Perché possiamo elevare $(1-p)$ alla $x-1$? In che senso “gli eventi che stiamo considerando sono indipendenti”? Cosa indichiamo con un evento? Le ripetizioni dell’esperimento di Bernoulli. Indichiamo con Y la variabile aleatoria descritta dal modello geometrico che conta i fallimenti prima del primo successo (contrapposta al modello geometrico che include il primo successo nella conta dei fallimenti); come cambia la funzione di massa? Qual è il valore atteso del modello geometrico? Calcoliamo il valore atteso di X , che segue la prima variante del modello geometrico: possiamo farlo a partire dal valore atteso di Y ? X ed Y sono così diverse? Quanto vale la differenza delle due? Una conta il numero di insuccessi più il primo successo, l’altra conta il numero di insuccessi: la differenza vale 1, quindi $X = Y + 1$; a partire da questo, e sapendo che $\mathcal{E}(X) = \frac{1-p}{p}$, posso calcolare il valore atteso di Y .

Quale proprietà del valore atteso stiamo sfruttando? Dimostriamo che il valore atteso di una costante è la costante stessa, avendo una variabile aleatoria discreta X , di cui conosco la funzione di massa, ed una funzione f : sono interessato al valore atteso di $f(X)$, cioè al valore atteso della nuova variabile aleatoria.

Quale proprietà di uno stimatore è catturata dalla consistenza? Come è definito lo scarto quadratico medio? Perché sapere che lo MSE tende a zero mi permette di dire che la varianza dello stimatore ed il bias tendono a zero? Perché sono entrambe due quantità positive.

Entropia. Come varia l’indice di Gini? Cosa succede alle mie frequenze quando ho un campione eterogeneo? Cosa sono n ed m ? Rispettivamente, la taglia del campione ed il numero di valori osservabili. Cosa succede all’indice di Gini in caso di massima eterogeneità?

Diagrammi di dispersione. Cosa deve succedere perché sussista una relazione di tipo diretto? La retta su cui si dovrebbero allineare i punti deve avere qualche proprietà particolare? Va bene una qualsiasi retta: infatti, posto che il coefficiente angolare è positivo, i punti di una qualunque retta associano un’ascissa grande ad un’ordinata grande. In quale diagramma cerchiamo espressamente di capire se i punti sono allineati sulla bisettrice del primo e del terzo quadrante? Un diagramma QQ. Nei diagrammi di dispersione ogni punto corrisponde a due valori accoppiati dei campioni; a cosa fanno riferimento i punti che disegno in un diagramma QQ? Cosa abbiamo sulle ascisse, e cosa sulle ordinate? Da una parte ho un campione, dall’altra una distribuzione: quindi sulle ascisse ho i quantili del campione, sulle ordinate quelli della distribuzione (o viceversa). Supponiamo di non voler standardizzare i dati: posso stimare dai dati che ho a disposizione i parametri della normale, cioè il valore atteso e la deviazione standard, con cui sto confrontando il campione? Sì, usando rispettivamente la media campionaria e la radice quadrata della varianza campionaria. Perché useremmo la media campionaria per stimare il valore atteso? Il valore atteso di una variabile aleatoria è un numero o un’altra variabile aleatoria? Come calcoliamo il valore atteso di una somma di variabili aleatorie?

Disegniamo il grafico della funzione di ripartizione di Bernoulli. Calcoliamo $P(X \leq 2)$, dove X è una Bernoulli. Quali valori può assumere X ? A cosa equivale calcolare la funzione di ripartizione nel punto x ? Qual è la probabilità che una Bernoulli sia minore o uguale di un numero negativo?