

Orali del 1 ottobre 2020

Alessandro Di Gioacchino

5 febbraio 2021

Probabilità condizionata. Perché la definizione cattura il concetto di “probabilità condizionata”? Che significato diamo al risultato? Ci indica come cambia la probabilità di A , sapendo che B è avvenuto (supponendo di voler calcolare $P(A | B)$)

Se gli eventi sono indipendenti, la probabilità della loro intersezione è uguale al prodotto delle singole probabilità. Grafico della funzione di ripartizione di un modello uniforme continuo nell'intervallo $[0, 10]$

Perché la definizione di probabilità condizionata ne cattura il concetto? Come modifica l'universo il fatto di sapere che si è verificato F ?

Assiomi di Kolmogorov. Quali semplici teoremi abbiamo dimostrato grazie a questi assiomi? (Qui basta parlare di come si calcola la probabilità del complemento di un evento)

Vogliamo stimare la varianza di una popolazione. Formalizzazione del problema della stima parametrica.

Per quale modello la varianza è un parametro? Poisson, oppure la normale (posto che il secondo parametro non sia la deviazione standard). La varianza, cioè quella grandezza che vogliamo stimare, coincide con Θ o con $\tau(\Theta)$? Sicuramente con Θ , ed in base alla distribuzione anche con $\tau(\Theta)$

Esiste un modello per cui la varianza campionaria è uno stimatore distorto? No, e lo abbiamo dimostrato. Consideriamo il seguente stimatore:

$$T := \bar{X}^2,$$

dove \bar{X} è la media campionaria della popolazione X . Può essere uno stimatore sensato per la varianza? In quali contesti? Se la varianza della popolazione è il quadrato del suo valore atteso, come può accadere nel modello esponenziale. Di solito la media campionaria è un buon stimatore per che cosa? Per il valore atteso, quindi il quadrato della media campionaria è un buon stimatore per il quadrato del valore atteso. Riusciamo a dire se tale stimatore è deviato o meno? Cosa deve accadere affinché il quadrato della media campionaria sia uno stimatore non deviato per la varianza? Sapendo che il valore atteso di $X \sim E(\lambda)$ è $\frac{1}{\lambda}$, posso in qualche modo dire che il valore atteso di \bar{X}^2 è $\frac{1}{\lambda^2}$? Potrei scrivere $\mathcal{E}(\bar{X}^2) = \mathcal{E}(\bar{X})\mathcal{E}(\bar{X})$? No, perché \bar{X} e \bar{X} non sono indipendenti (anzi, sono la stessa cosa). Il valore atteso è un operatore lineare, quindi posso portare fuori da esso solo scalature e traslazioni (**non** un elevamento a potenza): possiamo quindi dire che lo stimatore sia deviato? No, semplicemente non siamo in grado di rispondere con gli strumenti matematici presi in considerazione durante il corso.

Indice di Gini. La domanda non è chiara, perché ce ne sono due. Parliamo di quello per misurare la concentrazione. Come fa l'indice di Gini a capire dove mi trovo fra la situazione di concentrazione massima e quella di concentrazione minima? Misura l'area tra la bisettrice e la cosiddetta curva di Lorenz.

Abbiamo due variabili aleatorie indipendenti: quali proprietà possiamo derivare da questa?

Ad esempio, se X ed Y sono indipendenti, la funzione di ripartizione congiunta è uguale al prodotto delle marginali. Posso ricavare la funzione di massa di probabilità marginale a partire dalla congiunta indipendentemente dal fatto che X ed Y siano indipendenti.

Dimostriamo che $\mathcal{E}(XY) = \mathcal{E}(X)\mathcal{E}(Y)$ se X ed Y sono variabili aleatorie indipendenti, immaginando che X sia discreta ed Y continua. La sommatoria del valore atteso di una variabile aleatoria discreta è necessariamente una somma finita (cioè con n come secondo indice)? Ad ogni elemento della sommatoria corrisponde una specificazione della variabile aleatoria (e la probabilità che essa sia assunta); quindi dipende dal supporto della variabile aleatoria: nel caso della Poisson, poiché il supporto è \mathbb{N}^+ , la sommatoria non è finita.

Cosa cambia nel calcolo del valore atteso, passando da una variabile aleatoria discreta ad una continua? Scriviamo il valore atteso del prodotto di una variabile aleatoria discreta ed una continua; se fossero entrambe discrete o entrambe continue, potrei usare rispettivamente la funzione di massa o di densità congiunta. In questo caso dobbiamo invece usare una funzione in due variabili che, fissata X (quella discreta), è una funzione di densità e, fissata Y (quella continua), è una funzione di massa (non sono sicuro al 100% di questa cosa)

Scriviamo il valore atteso del prodotto di due variabili aleatorie discrete. Se X ed Y hanno un numero diverso di specificazioni, questa proprietà continua a valere? Sì, non cambia nulla.

Vogliamo stimare la varianza di una popolazione, con il seguente stimatore:

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

Lo stimatore non deviato per una quantità ignota è unico? Perché la correttezza è una proprietà desiderabile per uno stimatore? Se X è la popolazione ed X_i è il campione casuale estratto da X , X ed X_i seguono la stessa distribuzione: quindi posso usare X al posto di X_i , e viceversa.

In statistica descrittiva, per stimare la varianza di un campione usiamo lo stesso stimatore della statistica inferenziale, cioè la varianza campionaria. Come si comporta questo operatore se prende in input la trasformazione lineare di un campione? Sono costretto a calcolare daccapo la varianza? Perché è sensato che la traslazione non abbia impatto sul calcolo della varianza?

T^2 è veramente uno stimatore? Supponiamo di avere un campione che contiene il solo elemento 1: quanto vale T^2 ? $1 - \mu^2$: T^2 non è uno stimatore, perché dipende non solo dal campione ma anche da un parametro sconosciuto; se il parametro fosse conosciuto, T^2 sarebbe uno stimatore.

Come calcoliamo la varianza di un modello binomiale? (Non dare direttamente la formula, calcola la varianza usando una delle due definizioni a partire dal modello di Bernoulli)

Non basta l'indipendenza per dire che la varianza di una binomiale è n volte la varianza di una Bernoulli. Dovremmo invece esprimere formalmente X in base agli esiti degli n esperimenti di Bernoulli: X_1, \dots, X_n indicano l'esito della i -esima ripetizione. Che relazione c'è tra X e le n variabili aleatorie X_i ? Cosa vuol dire che X assume una certa specificazione, per esempio 3? La specificazione di una binomiale conteggia il numero di successi su n esperimenti di Bernoulli indipendenti, quindi $X = X_1 + \dots + X_n$

X assume valori da 0 ad n

Funzione di ripartizione di una Bernoulli. Evidenziamo nel grafico la probabilità che $X \leq \frac{1}{2}$ (non cambia nulla per $X < \frac{1}{2}$)

Evidenziamo nel grafico anche il valore atteso della variabile aleatoria (ricorda che è l'area compresa tra il grafico della funzione di ripartizione e la retta $y = 1$)

Abbiamo un campione di valori numerici, e vogliamo capire se questo è compatibile con una popolazione normale. Quali strumenti possiamo utilizzare? Come si chiama il diagramma? Come è individuato un punto all'interno del diagramma QQ? Cos'è un quantile? Che valori può assumere il 'parametro' di un quantile? Quali quantili prendiamo in considerazione per tracciare un diagramma QQ? Potremmo ad esempio prendere i cosiddetti 'decili' (cioè il quantile 0.1, il quantile 0.2, eccetera). Finora abbiamo parlato di quantili campionari: come estraiamo un quantile teorico da una distribuzione? Fissato q tra 0 ed 1, come calcolo il quantile di livello q della gaussiana? Cosa trovo sull'asse delle ascisse, se l'integrale da $-\infty$ ad x vale 0.3? Il quantile 0.3