Orali del 19 giugno 2020

Alessandro Di Gioacchino

5 febbraio 2021

A cosa serve un classificatore binario? Come faccio a valutarne la performance? Quale potrebbe essere una matrice di confusione ideale? Approssimativamente una diagonale, con tutti i valori che non sono sulla diagonale molto vicini a zero. Cosa usiamo se volessimo 'comprimere' la matrice in un numero più basso di valori? Cosa ci racconta la specificità? Esprime sempre le volte in cui il classificatore ha indovinato? La specificità, che ha veri positivi al numeratore e veri positivi più falsi negativi al denominatore, è la probabilità di quale evento? È la probabilità che, se il classificatore risponde "è positivo", l'oggetto in esame è effettivamente positivo. In altri termini, ci dice quale frazione di casi positivi il classificatore riesce ad indovinare? Di cosa tratta l'altro indice, la sensitività? Idealmente, sarei contento per quali valori di sensitività e specificità? Curva ROC. Perché il punto $(0, 1)$ indica il classificatore ideale? Cosa posso dire del numero di falsi positivi e del numero di falsi negativi se il classificatore si posiziona in tale punto? A cosa serve la diagonale nella grafico della curva? È possibile ottenere una curva degenere che giace su di essa? Sì, con un classificatore che sceglie in modo casuale.
Conosciamo solo il grafico della funzione di ripartizione di due v.a.
Disegna il grafico della funzione di ripartizione di una v.a. discreta. Cosa si intende con il simbolo di intervallo aperto all'inizio di ciascun tratto? Tracciamo nello stesso grafico la funzione di ripartizione di una v.a. continua, definita sull'intervallo da zero a tre (come le specificazioni della v.a. discreta di cui sopra). Anzi, riferiamoci ad una v.a. uniforme con specificazioni zero, uno, due. Guardando questo grafico, possiamo dire qualcosa sulla relazione tra i valori attesi delle due variabili aleatorie? Quella discreta potrebbe essere una binomiale? Quali specificazione ha? Una variabile con specificazioni uno, due, tre può essere una binomiale? No, dovrebbe esserci anche zero. Il ragionamento che dobbiamo fare coinvolge solo quanto il grafico esprime visivamente. Possiamo legare il grafico della funzione di ripartizione al valore atteso? Detto questo, siamo in grado di dire se c'è una particolare relazione tra il valore atteso della v.a. continua e della v.a. discreta? Sì: il valore atteso della v.a. discreta è maggiore.
Abbiamo un campione aleatorio descritto da n v.a., $X_1, \ldots, X_n$, e vogliamo stimare il valore atteso della popolazione. Normalmente cosa useremmo? Proviamo invece a stimarlo con la varianza campionaria. Qual è la sua definizione? Supponendo di avere già il valore della varianza campionaria, in quali situazioni usarlo per stimare il valore atteso potrebbe avere senso? Ce ne potrebbero essere alcune in cui tale stimatore non è neanche deviato, per esempio se la distribuzione ha valore atteso e varianza uguali. Quale distribuzione ha questa caratteristica? Se il valore atteso è invece funzione della varianza, lo stimatore potrebbe essere deviato: perché? Pensiamo all'esponenziale, in cui la varianza è il quadrato del valore atteso: in questo caso, per ottenere il valore atteso data la varianza campionaria, ne dovremmo estrarre la radice quadrata. Cosa possiamo dire circa le proprietà di questo stimatore? Cosa vuol dire che uno stimatore è corretto? Qual è il valore atteso di S, cioè della deviazione standard campionaria (radice della varianza campionaria)? $1/\lambda$
Sappiamo quindi che $\mathcal{E}[S] = \frac{1}{\lambda}$, perché S^2 è uno stimatore non deviato della varianza. Come indichiamo analiticamente questa cosa? Sappiamo che la varianza campionaria è sempre uno stimatore non deviato della varianza della popolazione. A partire dalla seconda equazione, $\mathcal{E}[S^2] = \frac{1}{\lambda^2}$, possiamo ricavare la prima? No, perché il valore atteso è lineare. Tornando al punto di partenza, possiamo quindi dire che la varianza campionaria è uno stimatore non deviato del valore atteso in quali casi? Quando la trasformazione che porta la varianza nel valore atteso è lineare, perché basta applicare la stessa trasformazione allo stimatore della varianza.

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 μ e σ 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 $\lim_{x\to\infty} F(x)$? 1
É 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 $X_1, \ldots, X_n$: 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 τ? Cos'è Θ? Il valore atteso di una popolazione può essere o meno un parametro: dipende dalla distribuzione. Per il momento indichiamo con μ 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 α 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 αX, 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: $P(\alpha X \leq y) = P\left(X \leq \frac{y}{\alpha}\right)$, perché $\alpha > 0$
A che probabilità ci siamo ricondotti? $P\left(X \leq \frac{y}{\alpha}\right) = 1 - e^{-\lambda \frac{y}{\alpha}}$
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 μ e σ? 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.

Orali del 11 settembre 2020

Alessandro Di Gioacchino

5 febbraio 2021

Cos'è un diagramma QQ? A cosa serve? Come lo generiamo? I dati si allineano su una qualunque retta o deve essere parallela alla bisettrice del primo-terzo quadrante? Perché il fatto che esista una relazione lineare tra il quantile del primo campione ed il quantile del secondo campione per ogni fissato livello avvalora l'ipotesi che i due campioni siano estratti dalla stessa distribuzione? Perché posso dire che la distribuzione è uguale solo se i quantili si allineano sulla bisettrice del primo-terzo quadrante? Se i percentili sono uguali indipendentemente dai valori, cosa posso dire?
Che differenza c'è tra combinazione e disposizione? Perché ne esistono due tipi ciascuna? Cosa cambia tra la versione con ripetizione e senza? A cosa serve il vincolo $k < n$? Come conteggio il numero di disposizioni di n oggetti su k posti con e senza ripetizione? Perché $\frac{n!}{(n-k)!}$ è il numero di disposizioni senza ripetizione di n oggetti in k posti? Perché calcolando $n!$ ottengo il numero di permutazioni di n oggetti? Quante scelte ho per l'ultimo oggetto? Solo una. Perché moltiplicando le possibili scelte ottengo il numero di permutazioni? Possiamo ragionare in questo stesso modo per calcolare il numero di disposizioni: al primo posto abbiamo n scelte, al secondo $n-1$, e così via fino al k-esimo dove ne ho $n-k+1$; la moltiplicazioni di questi valori coincide con $\frac{n!}{(n-k)!}$
Statistica inferenziale. Cosa vuol dire che uno stimatore è non distorto rispetto ad una quantità ignota? Perché è desiderabile che uno stimatore goda di questa proprietà? $\tau(\Theta)$ è la stima del parametro la quantità che voglio stimare? Perché ci sono Θ e $\tau(\Theta)$? Il primo è il parametro ignoto della distribuzione. Quando uno stimatore T è non deviato per Θ? (Ci stiamo momentaneamente dimenticando della funzione τ, c'è solo il parametro)
Le x in input allo stimatore sono variabili aleatorie o valori numerici? Cosa ottengo da questo calcolo? s, la stima della quantità. $\tau(\Theta)$ è quanto voglio stimare. Quindi ci aspettiamo che s sia una buona approssimazione per $\tau(\Theta)$

Cosa vuol dire che uno stimatore è non distorto? Perché ci piace che lo sia? Cosa indicano Θ e $\tau(\Theta)$? Esempio pratico. μ, il valore atteso di una popolazione, è un numero reale? Dimentichiamoci del fatto che esista τ per stimare solo Θ, quindi i valori dello stimatore sono stime per Θ: quando posso dire che lo stimatore T è non distorto? Formalmente come catturiamo il fatto che lo stimatore varia intorno a quanto vogliamo stimare? Supponiamo che la distribuzione sia un'esponenziale, e vogliamo sempre stimarne il valore atteso: cos'è Θ? Cos'è $\tau(\Theta)$? Θ è un parametro, cioè λ (in questo caso): siccome non sempre quanto vogliamo stimare è esattamente il valore di un parametro, ma una quantità che dipende da esso, il fatto di non conoscere quel parametro implica di non conoscere neanche quella quantità. Come troviamo uno stimatore per il valore atteso? Avrebbe senso usare la varianza campionaria per stimare il valore atteso di una popolazione esponenziale? Se calcolassimo il valore atteso dello stimatore, che risultato troveremmo supponendo che sia non distorto? È un calcolo che abbiamo già visto: quanto otteniamo come risultato? La varianza della popolazione. Cosa stima la varianza campionaria? 'Calcolare' la varianza campionaria significa "calcolare uno scarto medio:" ma tale scarto non è fatto rispetto al valore atteso delle singole variabili, perché nella varianza campionaria non c'è μ (cioè il valore atteso delle singole variabili) ma $\bar{X}$
Stiamo quindi calcolando una sorta di scarto medio rispetto al valore della media campionaria. Il numero che otteniamo in questo modo può essere usato per stimare qualcosa: cosa?
Distribuzione normale: aspetti rilevanti per descrivere questo modello. Il massimo della distribuzione rappresenta il valore atteso? Un punto è individuato da due valori, mentre il valore atteso è un numero solo (cioè l'ascissa del massimo). La regola empirica deriva dalla simmetria della distribuzione? No, è una proprietà tipica della normale. Una distribuzione normale può assumere valori negativi? Cosa individuano $\mu+\sigma$ e $\mu-\sigma$? I punti di flesso. Cerca di disegnare un buon grafico . . .
A cosa serve l'indice di correlazione? Stiamo parlando di statistica descrittiva, quindi cosa rappresentano x ed y? I campioni. Al denominatore abbiamo le deviazioni standard campionarie. Che tipo di relazione individua questo indice? Cos'è la covarianza? Ci permette di valutare la presenza di una relazione diretta o inversa: come ne interpretiamo il valore ottenuto? Come interpretiamo il valore ottenuto dall'indice di correlazione lineare? È vero che se è maggiore di zero la retta ha un coefficiente angolare positivo e se è minore di zero un coefficiente

angolare negativo, ma possiamo dire di più. Come è legata l'intercetta della retta (cioè q in $y = mx + q$) all'indice di correlazione? In nessun modo: la retta su cui si dispongono due campioni può toccare l'asse delle ordinate in un punto qualsiasi, sia se la relazione è lineare diretta che se la relazione è lineare inversa. L'indice di correlazione ha una proprietà interessante, che mi permette di interpretarlo in una maniera più efficace della covarianza. Quale valore ci aspettiamo di ottenere calcolando questo indice a partire da due campioni legati da un forte grado di relazione lineare diretta? È importante dire che i valori dell'indice di correlazione vanno tra -1 ed 1, perché in questo modo possiamo capire quanto sia forte l'eventuale relazione lineare.

Standardizzazione di variabili aleatorie: in cosa consiste? A cosa serve? Non c'è alcun valore, ma solo un modello; la standardizzazione non dipende dalle specificazioni, ma da qualcosa di più generale. Nel caso dei campioni, non riguarda il fatto di 'uniformare' la scala e non riguarda campioni diversi ma solo uno: effettuata la standardizzazione, il campione ha alcune proprietà interessanti.
Problema della stima parametrica e concetto di consistenza.

Calcoliamo la mediana di un modello esponenziale. Come si declina il concetto di mediana quando lo calcolo su una variabile aleatoria (o su una distribuzione)? Come definiamo la mediana di una variabile aleatoria? Conoscere il valore atteso di un'esponenziale potrebbe darmi una buona informazione sulla mediana? Come faremmo per calcolare la mediana di una distribuzione normale? È esattamente μ, ma la legge empirica non c'entra niente. Come calcoliamo il quantile di una distribuzione? Vediamo graficamente dove possiamo aspettarci di trovare la mediana dell'esponenziale (tracciando la funzione di densità). La distribuzione esponenziale mi permette di calcolare analiticamente il valore della mediana: come calcoliamo l'area sottesa dal grafico della funzione di densità tra 0 e la mediana? Usiamo la funzione di ripartizione. Cambia qualcosa se cominciassimo ad integrare da $-\infty$? A cosa è uguale l'integrale della funzione di densità da $-\infty$ alla mediana m? Alla funzione di ripartizione calcolata in m
Qual è la forma analitica della funzione di ripartizione di un'esponenziale? Se m è la mediana, quanto deve valere l'integrale calcolato? Poniamo il tutto uguale a $\frac{1}{2}$ e risolviamo rispetto ad m
Il logaritmo di una frazione è uguale alla differenza dei logaritmi, quindi $\ln \frac{1}{2} = -\ln 2$
Quindi la mediana di un'esponenziale è sempre $\frac{\ln 2}{\lambda}$
Confrontiamola con il valore atteso. Chi è più grande? $\ln 2$ è leggermente più piccolo di 1, quindi la mediana è più piccola del valore atteso.
Abbiamo un campione estratto da una popolazione esponenziale di parametro λ incognito: come potrei trovare uno stimatore per la mediana? Perché potremmo usare $\ln 2\bar{X}$ (dove X è la popolazione)? Se utilizzassi la media campionaria, otterrei $\frac{1}{\lambda}$ facendo che cosa? Calcolandone il valore atteso. Cosa preserva l'assenza di deviazione della media campionaria? La linearità del valore atteso: per capirlo basta applicare la definizione di stimatore non distorto. La linearità del valore atteso mi permette di portare fuori una costante moltiplicativa quale $\ln 2$
Poiché c'è una relazione lineare tra la mediana ed il valore atteso di un'esponenziale, posso sfruttare la linearità del valore atteso partendo da uno stimatore non deviato per il valore atteso della popolazione per ottenere uno stimatore non deviato per la mediana.
Quale altra proprietà degli stimatori abbiamo visto? Perché la consistenza è una proprietà desiderabile? Lo stimatore proposto gode di consistenza? Com'è definito lo scarto quadratico medio? Il motivo per cui la consistenza è una proprietà desiderabile diventa più chiaro usando la vera definizione dello scarto quadratico medio, e non la differenza tra varianza dello stimatore e bias rispetto alla quantità ignota. Il bias è una variabile aleatoria o un numero? Un numero, quindi non servirebbe calcolarne il valore atteso. Cosa calcola lo scarto quadratico? Calcolare il valore atteso della differenza tra quello che voglio stimare ed il valore atteso dello stimatore è una cosa; calcolare il valore atteso tra quello che voglio stimare e lo stimatore è un'altra.
Avendo dimostrato che lo stimatore è non distorto, come possiamo calcolare lo scarto quadratico medio? Come calcoliamo la varianza della media campionaria? Perché possiamo scrivere la varianza della somma delle X_i come la somma delle singole varianze? Perché sono indipendenti.
A cosa serve una curva ROC? Per visualizzare la bontà di un classificatore binario. Un classificatore fissato rispetto ad un campione individua un singolo punto nel grafico in cui tracciamo la curva ROC: dobbiamo prima introdurre il concetto di 'soglia.'

Variabile aleatoria uniforme continua. Quanto vale il grafico al di fuori dell'intervallo su cui la variabile aleatoria è definita? Vale zero. Calcoliamo il valore atteso di questa variabile. Perché quanto otteniamo è ragionevole? Qual è l'interpretazione geometrica del risultato ottenuto? Dove piazziamo $\frac{\alpha+\beta}{2}$ nel grafico di densità? (α e β sono gli estremi dell'intervallo in cui la variabile aleatoria è definita)
Perché la densità è costante, quindi ha senso che il baricentro dell'intervallo corrisponda al valore atteso.
Come stimiamo la mediana di una popolazione uniforme continua? Come calcoliamo la mediana della distribuzione

uniforme continua vista finora? A cosa è uguale? Al valore atteso, quindi usiamo la media campionaria.
Ho una popolazione descritta da una variabile aleatoria X distribuita secondo un modello uniforme di parametri α e β, entrambi ignoti; ho un campione casuale $X_1, \dots, X_n$ estratto da questa popolazione, e voglio trovare uno stimatore per la mediana della popolazione. Cosa si indica con Θ? Il parametro ignoto, quindi in questo caso sarebbe più la coppia (α, β)
A cosa è uguale la mediana? Che ragionamento si può fare per stimare la mediana? Cosa abbiamo scoperto di interessante circa la mediana di questa popolazione? Che è uguale al valore atteso: come ci può aiutare questa osservazione? Possiamo usare il valore atteso della distribuzione per stimare la mediana? No, perché è funzione di parametri sconosciuti quali α e β: in generale, come definiamo uno stimatore?
Abbiamo un campione numerico: quali strumenti possiamo utilizzare per capire se è stato estratto da una popolazione normale? Ragionando in termini di distribuzioni approssimativamente simmetriche, non basta richiedere che l'istogramma presenti una barra centrale più alta e sia simmetrico rispetto ad essa. Lo 'skew' indica una leggera asimmetria verso sinistra o destra. Bisogna imporre la presenza di una forma a campana, che è quanto accade se vale la regola empirica. Quale altro strumento grafico abbiamo visto, oltre all'istogramma? Il diagramma QQ: abbiamo un solo campione, e vogliamo capire se è compatibile con una popolazione normale. Tuttavia, per calcolare i quantili teorici della normale, devo conoscerne i parametri: come faccio?

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 stessimo 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 Θ 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)

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 'effe' 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 $P\left(X \leq \frac{\ln 2}{\lambda}\right) = \frac{1}{2}$: 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?

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 $P(A \mid B)$)
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: $X_1, \ldots, X_n$ 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 $X = X_1 + \cdots + X_n$
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

Orali del 2 ottobre 2020

Alessandro Di Gioacchino

5 febbraio 2021

Teorema di Bayes.
$P(A \mid B)$ vuole quantificare il rapporto di causalità tra A e B? No. Il teorema di Bayes stesso indica che non è stato B a scatenare A, perché considera anche $P(B \mid A)$

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Esercizio: ho un'urna con 10 palline, di cui 5 bianche e le altre nere. Estraiamo (senza re-immissione) due palline: E_1 è l'esito della prima estrazione, E_2 l'esito della seconda. Qual è la probabilità che E_2 sia bianca, dato che E_1 è bianca?
Stimatore non deviato per il parametro di un modello bernoulliano.

Abbiamo un campione in cui ogni osservazione è una coppia. Vogliamo calcolare le frequenze congiunte di questo campione. Dopo averle calcolate, come le organizzeremmo per mostrarle? La matrice in cui organizziamo le frequenze è di dimensioni $n \cdot n$, posto che n sia la taglia del campione? No, è una matrice $a \cdot b$ dove a sono i valori osservabili del primo carattere e b quelli del secondo. Cosa indica un generico elemento della matrice? Se indicasse la frequenza totale in cui queste coppie si ripetono, di che frequenza staremmo parlando? Assoluta. Che succede se sommo una colonna di quella matrice? Ottengo la frequenza marginale di un carattere della coppia; in altri termini, stiamo tenendo costante un elemento della coppia e facendo variare l'altro.
Immaginiamo di avere una coppia di variabili aleatorie. Il concetto di valore atteso si può estendere in modo da coinvolgerle entrambe: come?
Il valore atteso di una variabile aleatoria è aleatorio o costante? È costante. Un'espressione che coinvolge qualcosa di aleatorio ha come risultato qualcosa di aleatorio: quindi, siccome il valore atteso di una variabile aleatoria è costante, non compare alcuna variabile aleatoria nel calcolo di $\mathcal{E}(X+Y)$
Come calcoliamo $\mathcal{E}(X+Y)$
In quale caso possiamo scrivere la funzione di massa congiunta come prodotto delle marginali? Nel seguente passaggio, scriviamo la doppia sommatoria di una somma come la somma di due sommatorie:

$$\sum_x \sum_y (xP(X=x) + yP(Y=y)) = \sum_x \sum_y xP(X=x) + \sum_x \sum_y yP(Y=y)$$

Nel caso più generale possibile, cosa cambia se al posto della somma avessimo una generica funzione g?
Vogliamo stimare il valore atteso di una popolazione. Cos'è $\tau(\Theta)$? Una quantità ignota che dipende dal parametro Θ
Il valore atteso che vogliamo stimare è Θ o $\tau(\Theta)$? Sicuramente è $\tau(\Theta)$, eventualmente anche Θ (se il parametro della distribuzione è il valore atteso). Come stimiamo il valore atteso? Come ottengo la stima? Calcolando lo stimatore. Che stimatore utilizziamo di solito per il valore atteso? Com'è definita la media campionaria? Perché utilizziamo proprio lei? Dimostriamo che è uno stimatore non deviato per il valore atteso di una popolazione. Perché è desiderabile che uno stimatore sia non distorto? Cosa vuol dire che il valore atteso della popolazione è esattamente quanto voglio stimare? Cosa posso dire di una distribuzione, se conosco il suo valore atteso?

Orali del 5 ottobre 2020

Alessandro Di Gioacchino

5 febbraio 2021

(Estratto del primo orale, potrei aver interpretato male le domande)
Immaginiamo di avere un campione casuale estratto da una popolazione esponenziale: come ne stimiamo la mediana?
Se $X_1, \ldots, X_n$ indica il campione, cosa indica X? La popolazione. Nel problema della stima parametrica, Θ indica il parametro ignoto (che potrebbe non essere esattamente ciò che voglio stimare): non essendo la mediana un parametro della distribuzione, non posso usarlo come Θ. Supponiamo quindi che $\Theta = \lambda$: per stimare il parametro, possiamo usare il reciproco della media campionaria. Infatti sappiamo stimare 'facilmente' il valore atteso di una popolazione, ed il valore atteso di un'esponenziale è $\frac{1}{\lambda}$
Per stimare la mediana di un'esponenziale, che abbiamo dimostrato essere uguale a $\frac{\ln 2}{\lambda}$, è quindi ragionevole utilizzare $\ln 2\bar{X}$ come stimatore.

Torniamo alla popolazione esponenziale di cui sopra (il candidato non aveva risposto): se ne indichiamo la mediana con m, $m = \frac{\ln 2}{\lambda}$
Vogliamo stimare m, quindi $\tau(\Theta) = m$ e $\tau(\lambda) = \frac{\ln 2}{\lambda}$
La media campionaria è sempre uno stimatore consistente e non deviato per il valore atteso di una popolazione, ma ora vogliamo stimare λ (il parametro di un'esponenziale): non possiamo usare la media campionaria così com'è.
Dobbiamo fare un ragionamento che coinvolga il campione: riscriviamo il fatto che il valore atteso di un'esponenziale è $\frac{1}{\lambda}$ mettendo il campo la media campionaria (suppongo dovesse scrivere qualcosa del tipo: $\frac{1}{\lambda} = \mathcal{E}(\bar{X})$ In conclusione, se voglio stimare λ utilizzo il reciproco della media campionaria; se voglio stimare $m = \frac{\ln 2}{\lambda}$, utilizzo la media campionaria moltiplicata per $\ln 2$
Lo stimatore ottenuto è non deviato per la mediana? Il valore atteso della media campionaria è il valore atteso della popolazione, cioè $\frac{1}{\lambda}$; moltiplicando questo per $\ln 2$, otteniamo proprio quanto vogliamo stimare.
Curve ROC. L'estremo destro della curva è sempre il punto $(1, 1)$?

Abbiamo un'urna con 10 palline, di cui 4 nere e 6 bianche. Vogliamo contare il numero di palline bianche estratte. Quale distribuzione può modellare questa situazione? Potrebbero andar bene la ipergeometrica (estrazione senza re-immissione) o la binomiale (estrazione con re-immissione). Esistono altri modelli che posso applicare?
La distribuzione ipergeometrica è pensata proprio per situazioni di questo tipo. Supponiamo invece di reimmettere le palline una volta estratte: quali sono i parametri del modello binomiale risultante? Come calcoliamo la probabilità di non estrarre alcuna pallina bianca? Qual è la probabilità di estrarre almeno una pallina bianca?
Il coefficiente binomiale $\binom{n}{k}$ indica il numero di modi in cui posso costruire sottoinsiemi di k elementi a partire da n: se $k = 0$, il coefficiente binomiale vale 1 perché l'insieme vuoto è unico.
Dimostrazione: la probabilità dell'evento complementare è uguale ad $1-$ la probabilità dell'evento. Per dimostrarlo, usiamo gli assiomi di Kolmogorov in due punti: quando diciamo che la probabilità dell'evento certo è uno, e quando calcoliamo la probabilità dell'unione di eventi disgiunti come la somma delle singole probabilità.
Abbiamo una popolazione X, di cui conosciamo solo il valore atteso μ e la deviazione standard σ; vogliamo stimarne la varianza usando il seguente:

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

Quando uno stimatore si dice non deviato? Perché è una proprietà desiderabile?
È necessario elevare quella quantità al quadrato per calcolare il valore atteso di T^2? Cosa sappiamo di X_i? Fa parte di un gruppo di variabili aleatorie indipendenti ed identicamente distribuite. Posso dire che il valore atteso di X_i è μ? Perché sì? Perché seguono la stessa distribuzione della popolazione.
Cos'è uno stimatore? Uno stimatore è una funzione del solo campione; invece T^2 è anche funzione di μ: se μ è ignoto, T^2 non è uno stimatore.

Classificatori naive Bayes. Indichiamo con una variabile aleatoria Y l'esito della classificazione, e condizioniamo rispetto ad altre due variabili aleatorie, X_1 ed X_2, che rappresentano due attributi di interesse. Su quale ipotesi si basano questi classificatori 'ingenui?' Stiamo supponendo che gli eventi X_i (condizionati ad Y) siano indipendenti. C'è un altro punto di interesse, cioè non considerare il denominatore nel teorema di Bayes: perché? Cosa c'è di importante nel denominatore che mi permette di ignorarlo? Non dipende da k

Concetto di indipendenza: a cosa può riferirsi?
Gli assiomi di Kolmogorov sono sempre validi, per cui sappiamo che la probabilità dell'unione di eventi disgiunti è uguale alla somma delle singole probabilità; inoltre questo aspetto non c'entra con l'indipendenza. Cosa deve succedere perché due variabili aleatorie siano indipendenti?
Un evento si verifica; una variabile aleatoria no, però mi permette di ragionare in termini dell'evento "la variabile aleatoria ha assunto un certo valore?"
Che ruolo hanno x ed y, le specificazioni delle due variabili aleatorie? Possono assumere soltanto valori particolari? Non solo la definizione deve valere per ogni x ed y, ma in realtà si fa riferimento alla possibilità che i valori assunti dalle variabili aleatorie appartengano ad un certo insieme:

$$\forall A, B \quad P(X \in A, Y \in B) = P(X \in A)P(Y \in B) \iff X, Y \text{ indipendenti}$$

Quali interessanti proprietà possiamo ricavare sapendo che X ed Y sono indipendenti? Per esempio, se sono indipendenti la varianza della somma è uguale alla somma delle varianze: perché?
Dimostriamo che $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$
Ragioniamo sul concetto di indipendenza nell'ambito di statistica descrittiva. Come genero un diagramma di dispersione? Se prendo un generico punto del diagramma, cosa posso dire circa le sue coordinate? Sono le due componenti di una stessa coppia. Come calcoliamo il grado di correlazione dei due campioni? Se l'indice di correlazione campionaria è uno, i valori dei campioni si allineano su una qualsiasi retta. Cosa indica la presenza di una relazione diretta tra due quantità? Che al crescere di una quantità, cresce anche l'altra: per cui va bene ogni retta, purché abbia coefficiente angolare positivo.
Consistenza in statistica inferenziale. Lo scarto quadratico medio misura l'errore fatto stimando la quantità sconosciutà con lo stimatore: e il bias invece? In realtà, lo MSE si può scrivere come somma della varianza e del bias (quindi l'esaminando stava facendo un errore).

Concetto di indipendenza.
Se $P(E \cap F) = P(E)P(F)$, allora gli eventi E ed F sono indipendenti. In che modo questa definizione è legata alla probabilità condizionata?
Come estendiamo la definizione di indipendenza tra due eventi a tre eventi? Non basta imporre che siano indipendenti a coppie. Guarda il foglio #1
Cosa vuol dire che la media campionaria gode della proprietà di linearità?
Calcoliamo la varianza dello stimatore "media campionaria." È importante sottolineare che le conclusioni a cui si arrivano (in particolare che la stima migliora con il crescere del campione) sono legate al fatto che la media campionaria è uno stimatore non distorto per il valore atteso della popolazione.
Modello uniforme discreto. Grafico della funzione di ripartizione. Quanto vale questa funzione tra 0 e 1? Perché è ovvio che valga zero? Quali sono le specificazioni di questa variabile aleatoria? Ricorda sempre una variabile aleatoria discreta uniforme modella l'esito del lancio di un dado.

Varianza della media campionaria.
È sempre vero che la varianza di una somma è uguale alla somma della varianza? Perché la varianza delle singole X_i è uguale alla varianza della popolazione?
Quale teorema possiamo applicare per approssimare la distribuzione della media campionaria?
Per n che tende a $+\infty$, la distribuzione espressa dal teorema del limite centrale non è più approssimata ma esatta.
La distribuzione della media campionaria ottenuta tramite il teorema del limite centrale è coerente con il calcolo della sua varianza?
Eterogeneità.
Perché l'indice di Gini è compreso tra zero (incluso) ed uno (escluso)? L'indice vale zero nel caso di massima omogeneità: dimostrazione.
Un'applicazione dell'indice di Gini (ed in generale degli indici di eterogeneità). Gli alberi di decisione determinano la domanda da fare in base al valore dell'indice di Gini (o dell'entropia).

La variabile aleatoria Z è data dalla differenza tra due normali standard, X ed Y
Cosa possiamo dire del valore atteso di Z? Perché vale anch'esso zero? Non è vero che, siccome X ed Y sono

normali standard, allora anche Z lo è: il concetto di riproducibilità è legato alla famiglia della distribuzione, quindi al massimo potrei dire che anche Z è normale. Cosa c'è di strano nel dire che $Z \sim N(0,0)$? Cosa vuol dire che una variabile aleatoria ha varianza nulla? È sensato che la differenza di due normali standard dia origine ad una costante? Sarebbe come lanciare due dadi, sottrarre il risultato di uno al risultato dell'altro ed ottenere sempre lo stesso numero.

Calcoliamo la varianza di Z, ipotizzando pure che X ed Y siano indipendenti. Cosa succede alla varianza se $Z = X - 2Y$?

Quindi, $Z \sim N(0, \sqrt{2})$

Quale ulteriore modifica devo apportare perché Z sia una normale standard? La trasformazione si applica alla variabile aleatoria, non alle sue specificazioni. Basta quindi dividere Z per $\sqrt{2}$

Abbiamo una popolazione distribuita secondo il modello normale, con media 0 e deviazione standard ignota. Come posso stimare tale parametro? Di quali proprietà gode lo stimatore "deviazione standard campionaria"? È non deviato? Per dirlo, dovremmo calcolarne il valore atteso. Non abbiamo gli strumenti matematici adatti a farlo, anche perché il valore atteso non si può portare dentro una radice quadrata (la deviazione standard è la radice quadrata della varianza).

Curve ROC.

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 $\mathcal{E}[(X-\mu_X)(Y-\mu_Y)] = \mathcal{E}[X-\mu_X]E[Y-\mu_Y]$
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 approccia 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 $X_1, \ldots, X_n$ 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.