

La Stima Bayesiana 2 - 10/10 + 15/10 (Ricapitolazione e Metodi Monte Carlo)

Intro - 10/10

Nella scorsa lezione eravamo interessati alla stima della variabile aleatoria Y , avendo campioni di X . Volevamo effettuare la stima con approccio Bayesiano, considerato che diversamente da quando abbiamo fatto la verosimiglianza dovevamo stimare una variabile aleatoria e non un valore deterministico.



L'approccio classico si basa solo sui dati, sulle evidenze, allo scopo di trovare un parametro deterministico; invece l'approccio Bayesiano, sapendo che cerchiamo un parametro che non è deterministico ma che ha una distribuzione di probabilità, permette di inserire l'informazione su secondo me come dovrebbe essere fatta la media (info a priori) in un processo inferenziale che parte dai dati e fonde i dati e la conoscenza a priori per ottenere la conoscenza a posteriori.

Ricordiamo che nella scorsa lezione avevamo una Gaussiana a media 0 e varianza σ^2 .

Ma σ^2 è sempre noto? Proprio come in statistica classica la risposta è no. Il problema non sussiste, oltre a stimare la media bisognerà stimare anche la varianza, in pratica in Bayesiano avremo un modello prior per la media e un altro per la varianza, più in generale un vettore con tutte le cose da stimare.

Nel Bayesiano vogliamo trovare la posterior che sarà $f(Y|X)$, o θ invece di Y se ci sono più parametri.

Con più di due o tre dimensioni il Bayesiano è molto complesso quindi per anni è stato accantonato.

Sono poi nati degli algoritmi per risolvere il Bayesiano anche a moltissime dimensioni, integrare su centinaia di dimensioni resta impraticabile ma si può stimare la posterior non in maniera analitica ma invece in maniera campionaria, estraendo direttamente delle stime.

Più praticamente

$$\underline{\theta} \in \mathbb{R}^d \quad \underline{X}$$
$$\textcircled{H} \quad X \quad z \in \mathcal{U}_s$$

↳ theta minuscolo

θ è il parametro a più dimensioni sul quale vogliamo fare inferenza (ad una dimensione nell'esempio della scorsa lezione lo abbiamo chiamato Y).

X saranno i dati raccolti (in un vettore), funzione del parametro vettoriale θ del quale raccogliamo campioni e sul quale abbiamo informazioni a priori sulle sue singole dimensioni.

Cerchiamo la posterior:

$$p(\theta | x) = \dots$$

(nelle pdf al pedice si mette il nome delle variabili aleatorie, per questo le maiuscole)

Per trovare la posterior useremo sempre la versione semplificata della Regola di Bayes, cioè quella in cui c'è proporzionalità rispetto al denominatore.

La posteriori sarà quindi proporzionale a:

$$\propto p(\theta) p(x | \theta)$$

Annotations: $p(\theta)$ is prior, $p(x | \theta)$ is pdf, LIKELIHOOD, modello generativo.

Abbiamo detto che ad oggi la posterior si calcola in maniera numerica, non con un modello analitico ma invece attraverso l'estrazione dei campioni.



Gli algoritmi che effettuano il calcolo numerico della posterior prendono il nome di **Markov Chain Monte Carlo (MCMC)**.

Si tratta di particolari tecniche Monte Carlo a catena Markoviana, questi algoritmi convergono, facendo iterare un numero sufficientemente elevato di volte, alla distribuzione stazionaria che è proprio la posterior di nostro interesse. La caratteristica di MCMC è che l'estrazione successiva dipende dallo stato dell'ultima.

(MCMC)
permette di avere: θ_j a posteriori
come se $\{\theta_j\}_{j=1}^N$ generati da $p(\theta|x)$ quindi in accordo con la post
grazie a questo riusciamo ad applicare processi Bayesiani a più dimensioni

Le tecniche MCMC permettono quindi di ottenere in maniera campionaria la posterior, otteniamo i campioni di θ a posteriori, cioè i campioni di θ COME SE fossero generati da $f_{\theta|x}$, cioè in accordo alla posterior.



Grazie a questi valori di θ ottenuti è possibile calcolare, in maniera campionaria, i valori attesi, le pdf, le cdf, marginali e generali e in generale risolvere tutti i problemi di nostro interesse COME SE si fosse in possesso della $f_{\theta|x}$.

Si potrebbe risolvere il tutto analiticamente ma solo quando il problema è piccolo e magari "giocando" con le distribuzioni esce fuori una distribuzione nota e il problema è risolto ma questo non succede quasi mai e quindi occorrono degli algoritmi che numericamente permettono l'implementazione di tecniche Bayesiane anche in casi complessi.

Piccola ricapitolazione (penso da Ricapitolazione - 07/10 ad ora)

$\hat{\theta}$ con MMSE

Fino ad ora abbiamo stimato il parametro di nostro interesse che è stato prima Y (media della la posterior $f_{Y|X}(y|x)$ che era Gaussiana) e poi θ (uguale a Y ma in caso multidimensionale, quindi è un vettore, comunque sono solo nomi), tutto questo a sigma quadro fissato.

$$\hat{\theta}_{MMSE} = E[\theta | x] \quad \text{da integrale di } f(\theta|x)$$

media della gaussiana

Abbiamo stimato θ (in realtà Y monodimensionale ma il concetto è quello) con MMSE, cioè l'approccio Bayesiano, che è un approccio che trova $E[\theta|X]$ (minimizza l'errore quadratico medio), tutto questo ovviamente dopo aver raccolto i dati.

Ovviamente il valore atteso riguardava la pdf a posteriori e si calcola con l'integrale della distribuzione.

L'integrale della distribuzione era complesso ma era possibile risolvere il tutto analiticamente, carta e penna, con la Regola di Bayes e poi riconducendoci a distribuzioni note, questo non sempre è possibile e quindi poi servono algoritmi numerici.

$\hat{\theta}$ con MLE

Abbiamo anche stimato a massima verosimiglianza la θ che era la media della nostra Gaussiana.
(penso che ora sta riparlano del caso classico, è andato indietro)

$$\hat{\theta}_{ML} = \underset{\theta}{\operatorname{argmax}} f_{x|\theta}(x|\theta)$$

Questi sono due approcci ma ne esistono tanti altri.

$\hat{\theta}$ con MAP

Un altro approccio semplice da introdurre è **MAP, Maximum A Posterior**.

In altre parole si sceglie la $\hat{\theta}$ nel punto che massimizza la probabilità a posteriori, noi abbiamo una pdf e ci chiediamo qual è il valore massimo di questa pdf, cioè quello a maggiore probabilità; scegliamo $\hat{\theta}$ in accordo a questo.

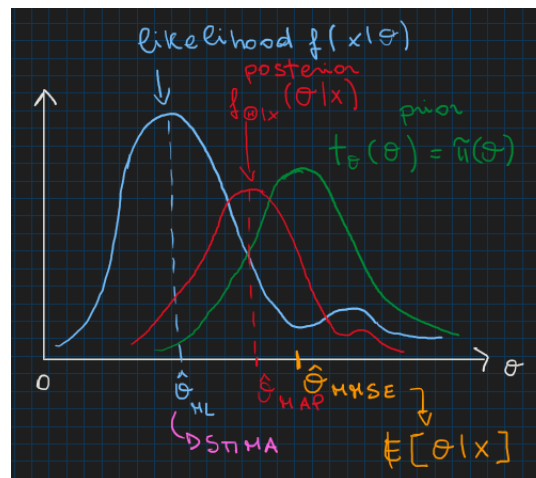
in aggiunta mostra il seguente

$$\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}} f_{\theta|x}(\theta|x)$$

↳ Maximum a posteriori

Questa regola quando è tutto Gaussiano coincide con la regola a minima distanza, serve per il progetto del decisore per il ricevitore digitale in Analisi dei Segnali. In quel caso si sceglie il simbolo che massimizza la probabilità a posteriori ma visto che tutti i simboli sono equiprobabili, nei casi semplici da noi studiati, diventa un approccio a massima verosimiglianza, l'approccio ML è quello che porta alla minima distanza nelle regioni di decisione.

Resoconto



Per la likelihood scegliamo $\hat{\theta}_{ML}$, che è quello che sull'asse delle θ corrisponde al valore massimo della likelihood.

Usando likelihood, blu, e prior, verde, vogliamo trovare la posterior, rossa.

Sappiamo che il prodotto tra prior e likelihood mi porta alla posterior a meno di una costante che serve a normalizzare.

Abbiamo detto che grazie alla **posterior** possiamo fare inferenza tramite l'**argmax**, questo ci porta ad avere $\hat{\theta}_{MAP}$.

L'MMSE invece essenzialmente "mediava" la posterior per mezzo dell'integrazione, il valore $\hat{\theta}_{MMSE}$ non è detto però che sia quello a massima probabilità (cioè MAP), ciò è verificato solo se la posterior è simmetrica.

Tornando a MCMC

Nel caso a più dimensioni, ma anche ad una sola dimensione se comunque il problema è troppo complesso da risolvere analiticamente, ci sono degli algoritmi che non permettono di calcolare la pdf MA permettono di estrarre campioni da essa, i campioni permettono poi di calcolare importanti grandezze come il valore atteso.

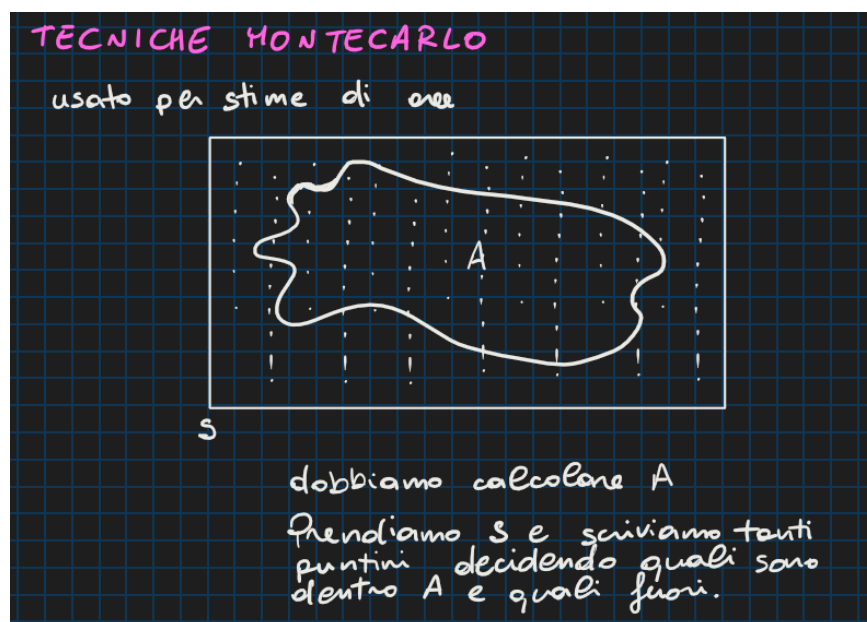
Il fatto che i campioni permettano di calcolare grandezze rilevanti era a noi già noto, quando abbiamo individuato le β che erano coefficienti della regressione lineare non avevamo la distribuzione dell'errore, abbiamo calcolato il tutto in modo campionario.

Tecniche Montecarlo - 15/10



Le tecniche Montecarlo nascono per poter calcolare oggetti complicati nei quali magari la parte probabilistica è troppo complicata e quindi si usano numeri casuali per assisterci nelle stime.

Un esempio per il calcolo di un'area aiuta a capire il concetto dietro le tecniche montecarlo.



Per fare l'area si può pensare all'integrale di una funzione se la forma è abbastanza trattabile, analiticamente facile, ma se questo non è vero o addirittura non abbiamo proprio la forma analitica la strategia dell'integrale non è viabile.

Però potrebbe essere che nonostante tutto possiamo tracciare l'area e quindi possiamo usare un approccio di tipo Montecarlo, cioè generare più volte dei numeri casuali e ripetere più volte l'esempio.

Partiamo da una superficie S della quale conosciamo le dimensioni e che contiene la superficie di interesse A .

A questo punto in maniera casuale si mettono puntini sulla superficie S e a questo punto possiamo contare quanti puntini sono dentro la superficie A e infine calcolare la superficie A come frazione dei punti al suo interno e dei punti totali.

Questo è un modo per calcolare le aree che non usa un approccio deterministico ma invece un approccio casuale basato sulla generazione randomica di punti.



Qualunque approccio basato sulla generazione di numeri casuali prende il nome di tecnica Montecarlo.

Il calcolo delle aree è solo un esempio di tecnica Montecarlo ma ce ne sono molte altre.

In R abbiamo stimato la percentuale di volte che la media valore vero è all'interno dell'intervallo di confidenza per verificare che davvero risulta essere $1-\alpha\%$.

IN BREVE

Ricordiamo in breve l'idea alla base del Metodo Monte Carlo: se dobbiamo stimare una quantità aleatoria e che dipenderà anche dalla variabilità campionaria ripetiamo, iteriamo, più volte l'estrazione di un campione casuale su cui applichiamo le stesse proprietà e poi andiamo a mediare; questo può essere usato per trovare stimatori e per capire mediamente il risultato restituito, o anche con che variabilità viene restituito il risultato (si può capire l'intervallo di confidenza capendo qual è la distribuzione di probabilità dei risultati e poi usando i quantili). Abbiamo fatto tutto questo in un esercizio in R.