

# **UNIVERSITÀ DEGLI STUDI DI SALERNO**



**Dipartimento di Ingegneria dell'Informazione ed  
Elettrica e Matematica applicata**

**Corso di Laurea in Ingegneria Informatica**

**APPUNTI DI DATA SCIENCE  
DI FRANCESCO PIO CIRILLO**

<https://github.com/francescopiocirillo>



"Sii sempre forte"

 Ehi, un attimo prima di iniziare!

Hai appena aperto una raccolta di appunti che ho deciso di condividere **gratuitamente** su GitHub, se ti sono utili fai **una buona azione digitale**:

-  **Lascia una stellina alla repo:** è gratis, indolore e fa super piacere!
-  **Condividerla con amici**, compagni di corso, o chiunque possa averne bisogno.

Insomma, se questi appunti ti salvano anche solo una giornata di studio... fammelo sapere con una **stellina!**

Grazie di cuore 

# La stima Bayesiana 3 - 17/10 (MSE stima parametro aleatorio Y)

## Valutiamo il MSE del nostro approccio per la stima del parametro aleatorio Y

Riprendiamo l'esercizio originale con il quale abbiamo introdotto la stima Bayesiana.

STIMA BAYESIANA

$$X = Y + W$$
$$\uparrow \quad \Leftrightarrow W \sim N(0, \sigma_w^2)$$
$$Y \sim N(0, \sigma_y^2) \quad \underline{\underline{E.Y}}$$

Ora vogliamo valutare l'errore dello stimatore calcolato con il nostro approccio che è il Mean Squared Error.



Prendiamo il Mean Squared Error e lo consideriamo come funzione di rischio che vogliamo ottimizzare, si parla infatti di Rischio Bayesiano.

Il Mean Squared Error è la media su X e Y della differenza al quadrato tra Y e la sua stima. La media è su X e Y perché ci sono due variabilità visto che può variare Y che è una variabile aleatoria ma può anche variare X, di nuovo variabile aleatoria, dal quale stiamo provando a fare inferenza.

Una media su due valori richiederebbe la congiunta su XY, per non doverla avere utilizziamo una proprietà dei valori attesi chiamata Tower Property o anche Legge dei Valori Attesi iterati, ha anche altri nomi.

Espressione generale:

$$\mathbb{E}_{x,y}[(y - \hat{y})^2] = \int_x \int_y (y - \hat{y}(x))^2 p(x, y) dy dx$$

calcoliamo il valore del MSE della stima  
 Theoretical error of MSE  
 $E_{x,y}[(Y - \hat{Y})^2] = E_y[E_x[(Y - \hat{Y})^2 | Y]]$   
 ↑  
 TOWER PROPERTY  $\Rightarrow E_{x,y}[h(x,y)] = E_y[E_x[h(x,y) | Y]]$   
 applichiamo la tower property allo stimatore Bayesiano

In pratica dividiamo la media sulle due variabili calcolando la media su Y della media su X della quantità di nostro interesse **condizionata** ad Y. Nell'applicazione della Tower Property si sceglie chi usare come condizionante a seconda di quali medie possono essere più facili da calcolare, non c'è una relazione d'ordine.

Dopo aver calcolato la media su X resterà una funzione su Y.

## La media interna

Partiamo dal calcolo della media interna ottenuta dalla Tower Property.

Ricordiamo il valore dello stimatore di Y:

Recall that est of Y is  $\hat{Y} = E[Y | X] = \alpha \bar{X} = \frac{\sigma_y^2}{\sigma_y^2 + \frac{\sigma_w^2}{N}} \bar{X}$

Iniziamo.

Inner exp value :  $E[(Y - \hat{Y})^2 | Y = y] =$   
 ↳ y piccolo perché dopo aver fissato Y la grande diventa piccola

Scriviamo  $Y = y$  perché visto che il valore atteso è su X concettualmente va calcolato per un fissato valore di Y, un valore fissato non è più maiuscolo ma è minuscolo.

$$= E[(Y - \alpha \bar{X})^2 | Y = y] = E[(y - \alpha(y + \frac{1}{N} \sum_i w_i))^2 | Y = y]$$

↓  
 diamo per conto che  $Y = y$  quindi non lo scriviamo più       $\bar{X} = y + \frac{1}{N} \sum_i w_i$   
 $\frac{1}{N} E[\cdot]$  →  $\sum_i w_i$

Sostituiamo il valore di  $\hat{Y}$  e poi sostituiamo anche il valore di  $\bar{X}$ . Invece di essere sommatoria di  $x_i$  è sommatoria di  $w_i$  con y sommato a parte perché y è stato fissato quindi è indipendente da i.

$$= E[(y - \alpha y - \frac{\alpha}{N} \sum_i w_i)^2] = E[(y(1-\alpha) - \frac{\alpha}{N} \sum_i w_i)^2]$$

Abbiamo ricondotto il contenuto della media al quadrato di una differenza che ora svolgiamo.

$$= \mathbb{E} \left[ y^2 (1-\alpha)^2 + \frac{\alpha^2}{N^2} (\sum_i w_i)^2 - 2y(1-\alpha) \frac{\alpha}{N} \sum_i w_i \right]$$

$\sum_i w_i^2 + \sum_i \sum_j w_i w_j$   
 valore atteso di questo

Avere un valore atteso con delle somme non è complicato, i quadrati potrebbero porre un problema.

Siccome alla fine bisognerà fare il Valore Atteso della somma dei Valori Attesi (questa media si divide in somme di medie ed è tutto contenuto nella media esterna).

Abbiamo diverse parti costanti che quindi sono semplici perché la media di una costante è la costante stessa **ma** ci resta il valore atteso della sommatoria al quadrato delle  $w_i$ , che sarà (in rosa) la sommatoria delle  $w_i^2$ , cioè tutti i quadrati, più tutti i doppi prodotti.

Di questo oggetto (rosa) è necessario calcolare il valore atteso.

$$\mathbb{E}[Z_1 Z_2] = \text{Cov}(Z_1, Z_2) + \mathbb{E}[Z_1] \cdot \mathbb{E}[Z_2]$$

$$\begin{aligned} & \sum_i w_i^2 + \sum_i \sum_j w_i w_j \\ & \text{valore atteso di questo} \\ & \text{cov}(z_1, z_2) = \mathbb{E}[(z_1 - \mathbb{E}[z_1])(z_2 - \mathbb{E}[z_2])] \\ & \stackrel{!}{=} \mathbb{E}[z_1] \mathbb{E}[z_2] = 0 \\ & \text{tornando in } w \\ & \mathbb{E}[\sum_i w_i^2] + \mathbb{E}[\sum_i \sum_j w_i w_j] \\ & \hookrightarrow N \text{Var}[w_i] \end{aligned}$$

della singola  $w$

La media del prodotto per variabili incorrelate equivale al prodotto delle medie, che in questo caso è 0.

La media della sommatoria sarà la sommatoria della media, la media della variabile aleatoria al quadrato è il momento del secondo ordine che per le variabili a media nulla equivale alla varianza.

Quindi in sintesi:

$$\begin{aligned}
&= \mathbb{E} \left[ y^2 (1-\alpha)^2 + \frac{\alpha^2}{N^2} \left( \sum_i w_i \right)^2 - 2y(1-\alpha) \frac{\alpha}{N} \sum_i w_i \right] \\
&\quad \text{SOMMA} \\
&\quad \text{di} \\
&\quad \text{valore atteso di questo} \\
&\quad \text{cov}(z_1, z_2) = \mathbb{E} \left[ (z_1 - \mathbb{E}[z_1])(z_2 - \mathbb{E}[z_2]) \right] \\
&\quad \text{tornando} \\
&\quad = \mathbb{E}[z_1] \mathbb{E}[z_2] = 0 \\
&\quad \mathbb{E} \left[ \sum_i w_i^2 \right] + \mathbb{E} \left[ \sum_i \sum_j w_i w_j \right] \\
&\quad \hookrightarrow N \text{ Var}[w_i] \\
&\quad \text{della singola } w \\
&= y^2 (1-\alpha)^2 + \frac{\alpha^2}{N} \sigma_w^2 + 0
\end{aligned}$$

Per quanto riguarda l'ultimo termine a destra, questo fa 0 perché è tutto moltiplicato per la somma delle medie delle  $w_i$  che sono tutte 0.

## La media esterna

ritorniamo in forma di  $y$  come variabile aleatoria

The outer expectation is :  $\mathbb{E}_y \left[ y^2 (1-\alpha)^2 + \frac{\alpha^2}{N} \sigma_w^2 \right] =$

Calcolata la media interna passiamo alla media esterna.

Per il calcolo della media interna avevamo "fissato"  $y$  che quindi era un valore costante, ora torniamo a considerare  $Y$  variabile aleatoria per calcolarne la media.

$$\begin{aligned}
&= (1-\alpha)^2 \mathbb{E}[y^2] + \frac{\alpha^2}{N} \sigma_w^2 = (1-\alpha)^2 \sigma_y^2 + \frac{\alpha^2 \sigma_w^2}{N} \\
\alpha &= \frac{\sigma_y^2}{\sigma_y^2 + \frac{\sigma_w^2}{N}}
\end{aligned}$$

Ancora una volta visto che la variabile aleatoria che stiamo considerando ha media 0 il suo momento del secondo ordine è pari alla varianza della variabile aleatoria, quindi  $\sigma_Y^2$ .

Ricordiamo che  $\alpha$  era un valore specifico che abbiamo definito precedentemente e che poteva essere diverso da 1.

## Risultato MSE

Sostituendo  $\alpha$  nel risultato che abbiamo ottenuto troviamo finalmente l'MSE.

$$\Rightarrow \mathbb{E}[(y - \hat{y})^2] = \frac{\sigma_y^2 \sigma_w^2}{N^2 (\sigma_y^2 + \sigma_w^2/N)^2} + \frac{\sigma_y^2 \sigma_w^2}{N (\sigma_y^2 + \sigma_w^2/N)^2} =$$

$$= \frac{\sigma_y^2 \sigma_w^2 (\sigma_w^2/N + \sigma_y^2)}{N (\sigma_y^2 + \sigma_w^2/N)^2} = \frac{\sigma_y^2 \sigma_w^2}{N \sigma_y^2 + \sigma_w^2} = \sigma^2$$

↳ l'abbiamo già trovata  
varianza di  $f(y|x)$



Notiamo che l'espressione finale che otteniamo per l'MSE è proprio pari alla varianza della posteriore  $f_{Y|X}(y|x)$ .

Quindi l'MSE è pari a quella che era la varianza dello stimatore, che poi è anche la varianza della posteriore, questo ha senso visto che nel Bayesiano il bias non ha valore l'MSE è di fatto determinato dalla varianza.

## Ricapitolazione

Se $\sigma_y^2 \rightarrow \infty$	$\mathbb{E}[(y - \hat{y})^2] = \frac{\sigma_w^2}{N}$	, MMSE = $\bar{x}$
$\sigma_y^2 \rightarrow 0$	" = 0	, MMSE = 0 = $\mu_y$
$\sigma_w^2 \rightarrow \infty$	" = $\sigma_y^2$	, MMSE = 0 = $\mu_y$
$\sigma_w^2 \rightarrow 0$	" = 0	, MMSE = $\bar{x}$

$$\sigma_y^2 \rightarrow \infty$$

Se la varianza della prior è infinita la prior non è informativa e quindi quello che domina l'MSE è la varianza dei dati, che però erano proprio dati dallo stimatore, il valore atteso di Y dato X, cioè l'MMSE.

L'MMSE con varianza infinita delle Y era proprio la  $\bar{x}$ .

Sto dicendo che senza informazioni a priori la varianza è  $\frac{\sigma_w^2}{N}$  cioè la stessa di quella dello stimatore a massima verosimiglianza, in altre parole solo i dati contano e nei dati la cosa importante è la W, cioè l'errore, perché è l'errore che crea la variabilità vera.

$$\sigma_y^2 \rightarrow 0$$

In questo caso il MMSE è 0 che è anche la media della Y.

Del resto se la prior è perfettamente affidabile ed informativa ha senso che il risultato che esce sia la media della distribuzione, che è 0.

$$\sigma_w^2 \rightarrow \infty$$

In questo caso resta solo  $\sigma_y^2$ .

Con una varianza enorme i dati sono inutili e solo la prior ha valore, è la varianza di  $y$  a dominare la varianza dello stimatore, l'MMSE è ancora una volta 0.

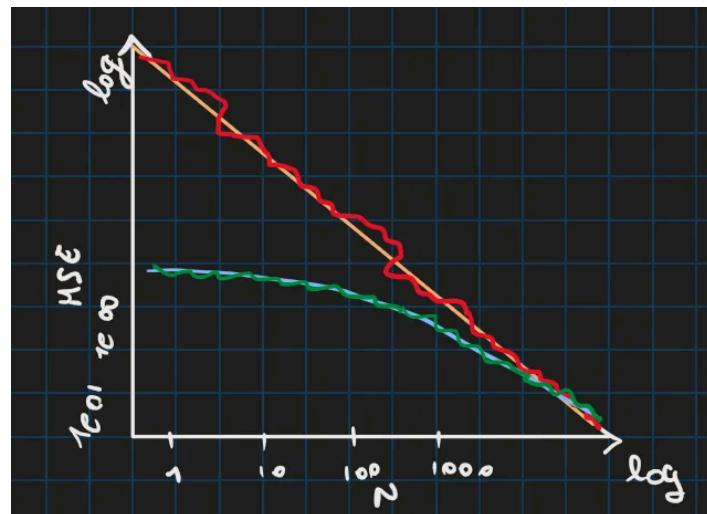
$$\sigma_w^2 \rightarrow 0$$

Se la varianza di  $W$  è 0 allora i dati hanno precisione assoluta quindi la media campionaria domina la stima.

## Conclusioni

In conclusione si riconferma quanto vedemmo nell'esercizio con i dati eterogenei, lo stimatore MMSE, cioè Bayesiano, si sposta dove c'è più precisione tra prior e dati.

## Importanza del grafico Log Log



Quella arancione è teorica la rossa è calcolata con Monte Carlo, lo stesso vale per blu e verde. Essenzialmente la varianza in MMSE è quella blu/verde, che tende ad essere quella di MLE ( $\sigma_w^2/N$ ) per  $N$  infinito che significa che i dati vincono sulla prior.

Un grafico Log Log è importante perché se abbiamo una relazione del tipo

$Y = x^\alpha$  se decidiamo di fare il logaritmo da entrambe le parti possiamo chiamarli poi  $y'$  e  $x'$ .

$$\log y' = \alpha \log x'$$

Quello che vediamo sono leggi di potenza, potenza significa anche -1 solo che c'è la potenza negativa.

Tutto questo è comodo perché  $x^b$  diventa  $b \cdot \log(x)$ , le cose diventano rette.



Il grafico Log Log permette di mettere in evidenza legami di tipo legge di potenza.

Siccome  $1/N$  è una legge di potenza in termini della nostra  $N$  che è la nostra  $X$ , il legame  $\sigma_w^2/N$  diventa una retta a pendenza -1, sarebbe sul grafico la linea arancione (teorica) e rossa (calcolata con Montecarlo).

La linea blu (teorica) e verde (calcolata con Montecarlo) è relativa all'MSE, questo parte dal  $\sigma_Y^2$ , sfrutta la priori all'inizio perché si ha un solo punto ma mano a mano che si va a destra il  $\sigma_w^2/N$  diventa sempre più piccolo e quindi si tende ad usare più i dati perché diventano più precisi del prior.

La legge è questa,  $\sigma^2/N$ , è una legge classica, quando gli stimatori seguono  $\sigma^2/N$  le deviazioni standard sono  $\sigma/\sqrt{N}$  che indica una decrescita di tipo  $\sqrt{N}$ , tendono al valore vero come  $\sqrt{N}$ , questi prendono il nome di radical N consistenti.

Se vanno più piano di radical N a 0 gli stimatori non sono molto apprezzati.

Chiudiamo con questo il tema dei Model Based.

Ora torniamo al concetto di regressione, con un approccio nel quale noi abbiamo sia  $X$  sia  $Y$  e costruiamo il rapporto, che non conosciamo trai due, approccio supervisionato.

## **DISCLAIMER**

Questi appunti sono stati realizzati a scopo puramente educativo e di condivisione della conoscenza. Non hanno alcun fine commerciale e non intendono violare alcun diritto d'autore o di proprietà intellettuale.

I contenuti di questo documento sono una rielaborazione personale di lezioni universitarie, materiali di studio e concetti appresi, espressi in modo originale ove possibile. Tuttavia, potrebbero includere riferimenti a fonti esterne, concetti accademici o traduzioni di materiale didattico fornito dai docenti o presente in libri di testo.

Se ritieni che questo documento contenga materiale di tua proprietà intellettuale e desideri richiederne la modifica o la rimozione, ti invito a contattarmi. Sarò disponibile a risolvere la questione nel minor tempo possibile.

In quanto autore di questi appunti non posso garantire l'accuratezza, la completezza o l'aggiornamento dei contenuti e non mi assumo alcuna responsabilità per eventuali errori, omissioni o danni derivanti dall'uso di queste informazioni. L'uso di questo materiale è a totale discrezione e responsabilità dell'utente.