Media di una Gaussiana con campioni non omoschedastici - 10/10 + Ricerca del MVUE - 14/10

Esercizio - stima della media di una Gaussiana a partire da campioni non omoschedastici - 10/10

Introduzione al problema

L'esercizio di oggi riprende a grandi linee la stima del parametro deterministico θ.

Da Ricapitolazione - 07/10 ad ora abbiamo sempre supposto σ^2 noto e omoschedasticità, ora proviamo qualcosa di diverso con σ^2 che cambia.

IMPOSTIAMO L'ESERCIZIO

sempre ipotesi di σ^2 dato e costante, ora vediamo cosa succede se varia

Usiamo per questo esercizio un approccio abbastanza semplice: MLE.

Useremo MLE per stimare la media di una Gaussiana (per rendere le cose semplici) ma ora abbiamo la caratteristica di avere dati eterogenei in termini di varianza.

MLE mean in dati eterogenei

$D = \{x_i\}_{i=1}^{N}$ $\quad \theta$ mean da stimare

$N/2$ var σ_1^2, $N/2$ var σ_2^2

modello del campione dal quale estraiamo i x_i

$$X_i \overset{iid}{\sim} N(\theta, \sigma_1^2) \quad i = 1, \dots, N/2$$

$$X_i \overset{iid}{\sim} N(\theta, \sigma_2^2) \quad i = \frac{N}{2}+1, \dots, N$$

Supponiamo N pari ma comunque è abbastanza irrilevante.

Questo modello con la varianza che a un certo punto cambia potrebbe sembrare assurdo ma in realtà è un modello ipersemplificato di qualcosa che può succedere. Banalmente possiamo immaginare che i dati sono oggetti che bisogna comprare, e ad esempio può essere che due negozi mi danno istanze di oggetti che fanno parte della stessa famiglia ma che da un negozio all'altro hanno varianza diversa. Può succedere e con numeri di cambiamenti di varianza sicuramente anche più alti di due.

Notiamo che il tutto equivale, ad eccezione della mancanza di omoschedasticità, all’esempio visto in Ricapitolazione - 07/10.

il tutto equivalente a

$$\Longleftrightarrow \quad X_i = \vartheta + W_i \qquad W_i \sim \mathcal{N}(0, \sigma_i^2) \quad \begin{array}{l} \sigma_i^2 = \sigma_1^2 \quad i = 1, \dots N/2 \\ \sigma_i^2 = \sigma_2^2 \quad i = \frac{N}{2}+1, \dots N \end{array}$$

Ovviamente visto il cambiamento di varianza le w saranno indipendenti ma non più identicamente distribuite.

Riflessioni iniziali

Se le varianze sono diverse possono emergere spontaneamente due domande alle quali si può rispondere studiando attentamente il problema:

- Visto che la varianza è indice di imprecisione in una stima non potrebbe essere conveniente prendere solo i dati a varianza più bassa e scartare gli altri?
  A
  posteriori si può decidere se buttare via dei dati può essere la scelta saggia, comunque in generale di solito non si buttano i dati, dicono sempre qualcosa.
- Potremmo mettere tutto insieme per l’MLE facendo la media di tutto e ignorando la varianza che varia?
  Per quanto riguarda la seconda domanda, facciamo la media e basta, tanto lo stimatore a massima verosimiglianza dice 1/n per somma delle
  x_i e la media è sempre la stessa quindi potremmo “ignorare” il problema della varianza che cambia… a quel punto però sorge la domanda: Questa è la migliore proposta?

Detto questo il nostro obiettivo diventa cercare di capire qual è la migliore proposta di soluzione per questo problema.

Calcolo della verosimiglianza

Grazie all'indipendenza sappiamo che la likelihood sarà data dal prodotto delle pdf e conosciamo le pdf grazie ai dati iniziali.

Si ricorda che la likelihood è la congiunta dei dati vista come funzione del parametro θ da stimare.

Visto che le pdf non sono tutte uguali ma sono di due tipi diversi invece di avere un’unica produttoria la likelihood sarà il prodotto di due produttorie.

$$L(\vartheta) = \prod_{i=1}^{N/2} \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{(x_i-\vartheta)^2}{2\sigma_1^2}\right\} \cdot \prod_{i=N/2}^{N} \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{-\frac{(x_i-\vartheta)^2}{2\sigma_2^2}\right\}$$

Le costanti come sempre possiamo eliminarle e a questo punto possiamo combinare gli esponenziali sommando gli esponenti.

likelihood $\Rightarrow$ congiunta dei dati vista come funzione del parametro θ da stimare

$$= \cancel{(2\pi\sigma_1^2)^{-\frac{N}{4}}} \; \cancel{(2\pi\sigma_2^2)^{-\frac{N}{4}}} \dots$$

non dipendono da θ

$$\propto \exp\left\{-\left[\frac{\sum_{i=1}^{N/2}(x_i-\theta)^2}{2\sigma_1^2} + \frac{\sum_{i=N/2+1}^{N}(x_i-\theta)^2}{2\sigma_2^2}\right]\right\}$$

A questo punto abbiamo ottenuto la funzione di verosimiglianza che una volta sostituite le x_i con i veri dati raccolti è una funzione di un solo parametro: θ.

Il punto in cui si massimizza la funzione ci da lo **stimatore a massima verosimiglianza**, se si sostituiscono le x_i troviamo la **stima a massima verosimiglianza**.

Abbiamo tolto le costanti, per massimizzare con maggiore comodità togliamo l'esponenziale calcolando la log likelihood e infine per rimuovere il segno meno ribaltiamo il problema e invece di calcolare l'argmax calcoliamo l'argmin.

$$\hat{\theta} = \arg\max_{\theta} \log L(\theta) = \arg\min_{\theta} \underbrace{\frac{\sum_{i=1}^{N/2}(x_i-\theta)^2}{2\sigma_1^2} + \frac{\sum_{i=N/2}^{N}(x_i-\theta)^2}{2\sigma_2^2}}_{g(\theta)}$$

Calcolo di $\hat{\theta}_{ML}$

Per trovare l'argmin si procede come sempre con le solite valutazioni sulle derivate. Iniziamo ponendo a 0 la derivata prima. La derivata seconda per aver trovato un punto di minimo dovrà essere positiva.

$$\frac{d}{d\theta} g(\theta) = 0 \iff \frac{1}{2\sigma_1^2} \cdot -2\sum_{i=1}^{N/2}(x_i-\theta) + \frac{1}{2\sigma_2^2} \cdot -2\sum_{i=N/2+1}^{N}(x_i-\theta) = 0$$

$$\iff \frac{1}{\sigma_1^2}\sum_{i=1}^{N/2} x_i - \frac{1}{\sigma_1^2}\frac{N}{2}\theta + \frac{1}{\sigma_2^2}\sum_{i=\frac{N}{2}+1}^{N} x_i - \frac{1}{\sigma_2^2}\frac{N}{2}\theta = 0$$

$$\iff \frac{1}{\sigma_1^2}\theta + \frac{1}{\sigma_2^2}\theta = \frac{1}{\sigma_1^2}\underbrace{\frac{\sum_{i=1}^{N/2} x_i}{N/2}}_{\bar{x}_1} + \frac{1}{\sigma_2^2}\underbrace{\frac{\sum_{i=N/2+1}^{N} x_i}{N/2}}_{\bar{x}_2}$$

Il primo rigo è la derivata.

Al secondo rigo abbiamo diviso le sommatorie.

Al terzo rigo portiamo gli elementi negativi a destra e diventano così positivi, inoltre dividiamo tutto per N/2.

Notiamo che le parti in rosa sono le medie campionarie di x_1 e x_2.

A questo punto effettuiamo le somme a sinistra e a destra e otteniamo:

$$\Leftrightarrow \theta \frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2} = \frac{\sigma_2^2\bar{x}_1+\sigma_1^2\bar{x}_2}{\sigma_1^2\sigma_2^2}$$

Isolando θ troviamo l'espressione per il calcolo di $\hat{\theta}_{ML}$.

$$\Leftrightarrow \hat{\theta}_{ML} = \frac{\sigma_2^2\bar{x}_1+\sigma_1^2\bar{x}_2}{\sigma_1^2+\sigma_2^2} \Leftrightarrow$$

$$\Leftrightarrow \boxed{\hat{\theta}_{ML} = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\bar{x}_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\bar{x}_2}$$

media pesata delle due medie
dipende dalle sole varianze

Considerazioni finali

> 💡 Abbiamo scoperto che la stima a massima verosimiglianza della media della Gaussiana della quale abbiamo campioni con due diverse varianze sarà la media pesata delle medie campionarie delle due parti a diversa varianza.
> I pesi di questa somma dipendono proprio dalle diverse varianze in gioco, nello specifico sono la varianza dell'altro fratto la somma delle varianze.

Cosa succede quando le due varianze sono diverse?

Visto che stiamo considerando una media pesata dove i due pesi sono frazioni con denominatore uguale ciò che è importante sono i numeratori.

SE $\sigma_2^2 > \sigma_1^2$ ha più peso la prima quindi quella che è più precisa

Se $\sigma_2^2 > \sigma_1^2$ allora avrà più peso la media campionaria dei primi campioni, il che ha senso se si pensa che i primi campioni avendo varianza minore sono quelli più affidabili.

Quindi capiamo che la strategia migliore è combinare i campioni ma pesandoli opportunamente con le varianze in modo da dare più enfasi a quelli più precisi.

Il risultato specifico che abbiamo ottenuto è basato sul fatto che abbiamo tutto Gaussiano ma comunque il concetto delle somme pesate che abbiamo individuato è sempre valido.

Se non avessimo avuto le varianze?

Se non avessimo avuto le varianze avremmo prima dovuto trovare le medie campionarie poi usarle per la stima delle varianze e poi arrivare a questo punto. Con un approccio plug-in inserivamo al posto delle varianze le stime delle varianze.

Se non fossero state Gaussiane?

Se non fossero state Gaussiane cambiavano le forme funzionali ma comunque asintoticamente nelle verosimiglianze in genere compaiono spesso queste medie, "sommatorie di…", quando compaiono medie e sommatorie alla fine in qualche modo per N sufficientemente grande si torna agli approcci Gaussiani.

Nelle valutazioni delle prestazioni spesso si decide di trattare il problema come se lo stimatore MLE fosse Gaussiano, perché con N che va a infinito e presenza di medie e sommatorie ci viene in aiuto il Teorema del Limite Centrale.

Anche se le mie X_i fossero debolmente correlate (noi lo enunciammo nel caso tutto iid e omoschedastico) comunque ci aiuterebbe il TLC perché esistono tantissime generalizzazioni che permettono di usare sempre il TLC e alla fine l'unica cosa che cambia è la velocità di convergenza.

A quanto equivale la somma dei coefficienti?

La somma dei due coefficienti è 1.

Chiamiamo il primo coefficiente q e otteniamo quanto segue:

somma i due coef $= 1$

$$q \triangleq \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \rightarrow \boxed{\hat{\theta}_{ML} = q\bar{x}_1 + (1-q)\bar{x}_2}$$

> 💡 Lo stimatore ML è quindi la **combinazione convessa** delle medie campionarie.

Si parla di combinazione convessa quando la combinazione lineare presenta coefficienti la cui somma è 1.

La cosa interessante delle combinazioni convesse è che al variare di q qualunque elemento resta incluso in un intervallo che in questo caso è tra $\overline{x_1}$ e $\overline{x_2}$.

COMBINAZIONE CONVESSA

↳ definiscono un intervallo

Troveremo spesso le Combinazioni Convesse, del resto le cose che sommate fanno 1 ricordano le PMF.

> 💡 Il coefficiente sarà più alto per il dataset a maggior precisione.

Casi particolari

Primo caso particolare

a) $\sigma_1^2 = \sigma_2^2 \Rightarrow q = \frac{1}{2}$

$$\hat{\theta}_{ML} = \frac{1}{2}\overline{X_1} + \frac{1}{2}\overline{X_2} = \frac{1}{\cancel{2}} \frac{\sum_{i=1}^{N/2} x_i}{N/\cancel{2}} + \frac{1}{\cancel{2}} \frac{\sum_{N/2+1}^{N} x_i}{N/\cancel{2}} = \frac{\sum_{i=1}^{N} x_i}{N} = \overline{X}$$

Quindi questa formula con la combinazione convessa se c'è varianza costante restituisce, giustamente, lo stimatore a massima verosimiglianza.

> 💡 Una proposta di stimatore, anche se le due varianze sono diverse, potrebbe essere la media dello stimatore media campionaria sulla prima parte e dello stimatore media campionaria sulla seconda parte.

— media tra le varianze —

$$\hat{\theta}_{avg} = \frac{1}{2}\overline{X_1} + \frac{1}{2}\overline{X_2} = \overline{X}$$

Questo corrisponderà allo stimatore a massima verosimiglianza SOLO quando la varianza è costante.

Questo stimatore, $\hat{\theta}_{AVG}$, indipendentemente dalle σ^2 in gioco vuole stimare la media con la media campionaria.

Intendiamo confrontare lo stimatore a massima verosimiglianza con lo stimatore che prende tutti i dati e ne fa la media campionaria senza valutare le differenze di varianza.

Secondo caso particolare

b) $\sigma_2^2 >> \sigma_1^2 \Rightarrow q \simeq 1 \Rightarrow \hat{\theta}_{ML} \simeq \overline{x_1}$

però perdo molta precisione quindi nei casi reali non ignorerò mai dei dati

Se c'è una differenza così grande di varianza, per σ_2^2 che tende ad infinito, allora conviene scartare dei dati MA nei casi pratici praticamente mai si scartano dei dati.

Terzo caso particolare

c) $\sigma_1^2 >> \sigma_2^2 \Rightarrow q \simeq 0 \Rightarrow \hat{\theta}_{ML} \simeq \overline{x_2}$

Valgono le stesse considerazioni del caso precedente.

Confronto tra diverse proposte di stimatori

Abbiamo per ora 3 proposte:

$$\hat{\theta}_{ML}, \quad \hat{\theta}_{avg}, \quad \hat{\theta}_i = \overline{x_i} \quad i = 1, 2$$

- $\hat{\theta}_{ML}$ che è il più tradizionale che abbiamo anche ricavato con i calcoli;
- $\hat{\theta}_{AVG}$ che è quello che ignora il fatto che la varianza cambia;
- $\hat{\theta}_i$ che è lo stimatore basato solo su una parte dei dati che hanno la stessa varianza (ad esempio possiamo considerare solo i dati a varianza σ_1^2 e ignorare gli altri).

La prima domanda che bisogna chiedersi è se questi stimatori hanno senso e se hanno senso come facciamo a confrontarli in termini di efficienza e prestazioni?

Un primo criterio importante è la **unbiasedness**.

$$\mathbb{E}[\hat{\theta}_1] = \mathbb{E}\left[\frac{1}{N/2}\sum_{i=1}^{N/2} x_i\right] = \frac{2}{N}\sum_{i=1}^{N/2} \underbrace{\mathbb{E}[x_i]}_{\theta} = \theta$$

$$\mathbb{E}[\hat{\theta}_2] = \ldots \quad \text{uguale} \uparrow \quad = \theta$$

$$\mathbb{E}[\hat{\theta}_{avg}] = \mathbb{E}[\overline{x}] = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}[x_i] = \theta$$

$$\mathbb{E}[\hat{\theta}_{ML}] = \mathbb{E}[q\,\overline{x}_1 + (1-q)\,\overline{x}_2] = q\,\mathbb{E}[\overline{x}_1] + (1-q)\,\mathbb{E}[\overline{x}_2] = q\theta + (1-q)\theta = \theta$$

Tutti gli stimatori in questione risultano unbiased.

Visto che sono tutti a bias 0 per confrontarli sarà di interesse il calcolo della varianza.

Il migliore sarà il **Minimum Variance Estimator**.

Visto che sono tutti unbiased si parlerà di **Minimum Variance Unbiased Estimator** o MVUE.

Ricerca del MVUE - 14/10

Nella statistica classica le metriche di riferimento per la valutazione della qualità di uno stimatore erano due: bias e varianza; che insieme vanno a costruire l'unica metrica tenuta in considerazione oggi cioè il Mean Squared Error (che è proprio $var + bias^2$).

Nella statistica tradizionale la ricerca degli stimatori era sempre fatta tra quelli unbiased e noi abbiamo seguito l'approccio classico, prima abbiamo valutato la correttezza degli stimatori e ora ci interessa la varianza.

Giunti alla fine della scorsa lezione alla conclusione che $\hat{\theta}_{ML}$, $\hat{\theta}_{AVG}$ e $\hat{\theta}_i$ sono tutti unbiased la ricerca dello stimatore migliore diventa la ricerca tra questi dello stimatore a minima varianza, questo prenderà il nome di Minimum Variance Unbiased Estimator.

Una volta stabilito che tutti gli stimatori sono unbiased il MSE equivale alla varianza.

💡 In genere gli stimatori a varianza minima sono detti efficienti.

Come detto in passato esiste il **Cramér–Rao bound** per il quale non si potrà mai scendere sotto una certa varianza, l'MLE arriva al bound asintoticamente.

Calcolo delle varianze

$\hat{\theta}_i$

$$\mathrm{Var}[\hat{\theta}_1] = \mathrm{var}[\bar{x}_1] = \mathrm{var}\left[\frac{2}{N}\sum_{1}^{N/2} x_i\right] \overset{\text{ind} \Rightarrow \text{incorrelati}}{=} \frac{4}{N^2} \cdot \frac{N}{2}\sigma_1^2 = \frac{2}{N}\sigma_1^2$$

$$\mathrm{Var}[\hat{\theta}_2] = \mathrm{var}[\bar{x}_2] = \frac{2}{N}\sigma_2^2$$

$\hat{\theta}_{AVG}$

$$\mathrm{Var}[\hat{\theta}_{avg}] = \mathrm{var}[\bar{x}] = \mathrm{var}\left[\frac{1}{2}\bar{x}_1 + \frac{1}{2}\bar{x}_2\right] =$$

$$\frac{1}{4}\mathrm{var}[\bar{x}_1 + \bar{x}_2] \underset{\text{incorrelazione}}{=} \frac{1}{4}\cdot\frac{2}{N}\sigma_1^2 + \frac{1}{4}\frac{2}{N}\sigma_2^2 = \frac{\sigma_1^2 + \sigma_2^2}{2N}$$

$\hat{\theta}_{ML}$

$$\mathrm{Var}[\hat{\theta}_{ML}] = \mathrm{Var}[q\bar{x}_1 + (1-q)\bar{x}_2] = \qquad q \triangleq \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$$

$$= \frac{\sigma_2^4}{(\sigma_1^2+\sigma_2^2)^2}\cdot\frac{2}{N}\sigma_1^2 + \frac{\sigma_1^4}{(\sigma_1^2+\sigma_2^2)^2}\cdot\frac{2}{N}\sigma_2^2$$

$$= \frac{2}{N}\frac{\sigma_1^2\sigma_2^4 + \sigma_1^4\sigma_2^2}{(\sigma_1^2+\sigma_2^2)^2} = \frac{2}{N}\sigma_1^2\sigma_2^2\frac{\cancel{\sigma_2^2+\sigma_1^2}}{(\sigma_1^2+\sigma_2^2)^{\cancel{2}}}$$

$$= \frac{2}{N}\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} = \frac{1}{N}\frac{2}{\frac{1}{\sigma_2^2}+\frac{1}{\sigma_1^2}} \quad \xrightarrow{\sigma_1^2\to\infty} \frac{\sigma_2^2}{N/2}, \quad \xrightarrow{\sigma_2^2\to\infty} \frac{\sigma_1^2}{N/2}$$

→ stimatore della var nella parte dei dati più precisi

Notiamo che continua ad esserci una somiglianza con una media armonica.

Notiamo inoltre che se una delle due varianze tende ad infinito lo stimatore ML converge allo stimatore della varianza sulla parte dei dati più precisa.

Questo rafforza che in generale non conviene eliminare i dati, perché solo asintoticamente, in casi estremi che non esistono, lo stimatore ML diventa quello relativo solo ad una parte dei dati.

Confronti

$var[\hat{\theta}_{ML}] < var[\hat{\theta}_i]$

i) $\mathrm{var}[\hat{\theta}_{ML}] < \mathrm{var}[\hat{\theta}_1]$

Pf. $\frac{2}{N}\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} < \frac{2}{N}\sigma_1^2 \iff \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2} < 1$ True

Dai calcoli risulta una disuguaglianza sempre verificata, quindi risulta **sempre vero** che lo stimatore a massima verosimiglianza ha varianza minore dello stimatore per una specifica parte del dataset.

$$var[\hat{\theta}_{ML}] < var[\hat{\theta}_{AVG}]$$

Che lo stimatore ML, che usa tutti i campioni, fosse migliore di uno stimatore che non usa tutti i campioni ma solo una parte si poteva dedurre; un confronto più interessante potrebbe essere con un altro stimatore che usa tutti i campioni, cioè $\hat{\theta}_{AVG}$, che comunque ricordiamo sacrifica una parte dell'informazione nel senso che non pesa la somma delle medie campionarie considerando la differenza delle varianze.

ii) $var[\hat{\theta}_{ML}] < var[\hat{\theta}_{avg}]$

Pf. $\frac{2}{N}\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} < \frac{\sigma_1^2+\sigma_2^2}{2N} \iff 4\sigma_1^2\sigma_2^2 < (\sigma_1^2+\sigma_2^2)^2$

$\iff 4\sigma_1^2\sigma_2^2 < \sigma_1^4 + 2\sigma_1^2\sigma_2^2 + \sigma_2^4 \iff (\sigma_1^2-\sigma_2^2)^2 > 0$ True if $\sigma_1^2 \neq \sigma_2^2$

If $\sigma_1^2 = \sigma_2^2$ $\quad \hat{\theta}_{ML} = \hat{\theta}_{avg}$

Dai calcoli si ricava che la disuguaglianza è sempre vera tranne quando le varianze sono uguali, in quel caso lo stimatore ML e lo stimatore della media campionaria coincidono.

$$var[\hat{\theta}_{AVG}] < var[\hat{\theta}_i]$$

Concettualmente possiamo immaginare che usare tutti i campioni invece di una parte possa condurre a maggiore forza nella stima.

iii) $var[\hat{\theta}_{avg}] < var[\hat{\theta}_1]$

$var[\bar{x}] < var[\bar{x}_1] \iff \frac{\sigma_1^2+\sigma_2^2}{2\cancel{N}} < \frac{\sigma_1^2}{\cancel{N}/2}$

$\iff \frac{\sigma_1^2+\sigma_2^2}{4} < \sigma_1^2 \iff \sigma_2^2 < 3\sigma_1^2$

opportuna combinazione convessa
$\Downarrow$
minimizza la varianza

Scopriamo che ciò è vero quando la varianza della seconda parte è minore di 3 volte rispetto alla varianza della prima, altrimenti si ribalta la relazione d'ordine.

Ovviamente resta vero che la ML è migliore, fa la combinazione convessa opportuna tenendo conto con precisione dei pesi stabiliti dalle varianze.

Ma nell'ambito delle combinazioni convesse si può addirittura migliorare il risultato rispetto a quello della specifica combinazione convessa stabilita da MLE?

La famiglia di stimatori $\hat{\theta}_{ALL}$

I precedenti stimatori sono tutti derivabili dalla seguente espressione:

$$\hat{\theta}_{all} = \delta \bar{x}_1 + (1-\delta)\bar{x}_2 \qquad \delta \in [0,1]$$

↳ TUTTI appartengono a questa famiglia

I 4 stimatori precedentemente visti sono:

$$\hat{\theta}_1 = \hat{\theta}_{all} \quad \text{per} \quad \delta = 1$$
$$\hat{\theta}_2 = \hat{\theta}_{all} \quad \text{per} \quad \delta = \emptyset$$
$$\hat{\theta}_{avg} = \hat{\theta}_{all} \quad \text{per} \quad \delta = \frac{1}{2}$$
$$\hat{\theta}_{ML} = \hat{\theta}_{all} \quad \text{per} \quad \delta = q \triangleq \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

q è quel rapporto speciale che abbiamo precedentemente trovato.

Quindi ci poniamo il quesito: qual è lo stimatore della famiglia $\hat{\theta}_{ALL}$ che minimizza la varianza? Può essere che esiste una combinazione convessa che modificando un po' la q che abbiamo trovato è addirittura migliore della ML?

Calcoliamo la varianza di $\hat{\theta}_{ALL}$:

quale di questi stimatori minimizza la varianza?

$$\text{var}[\hat{\theta}_{all}] = \delta^2 \text{Var}[\bar{x}_1] + (1-\delta)^2 \text{Var}[\bar{x}_2] \overset{(*)}{=} \delta^2 \frac{\sigma_1^2}{N/2} + (1-\delta)^2 \frac{\sigma_2^2}{N/2}$$

A questo punto abbiamo ottenuto la varianza come funzione del parametro δ e noi vogliamo trovare il valore di δ che minimizza questa funzione.

Find $\delta^* = \arg\min f(\delta)$

$$(**) \quad \frac{d}{d\delta} f(\delta) \overset{(*)}{=} \cancel{2}\,\delta \frac{\sigma_1^2}{\cancel{N/2}} - \cancel{2}(1-\delta)\frac{\sigma_2^2}{\cancel{N/2}} = 0 \Rightarrow \delta(\sigma_1^2 + \sigma_2^2) = \sigma_2^2$$

$$\Rightarrow \delta^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = q \Rightarrow \text{ML è il migliore di questa fam.}$$

$$\hat{\theta}_{all} = \hat{\theta}_{ML} \quad \text{MLE}$$

$$\text{From } (**) \rightarrow f''(\delta) = \frac{4}{N}\sigma_1^2 + \frac{4}{N}\sigma_2^2 > 0$$

Anche facendo la derivata seconda risulta che abbiamo effettivamente trovato un minimo in quanto è positiva.

Dai calcoli risulta che la δ^* è proprio q e quindi lo stimatore che minimizza la varianza sarà sempre maximum likelihood (nella famiglia degli stimatori con la combinazione convessa).