Capitolo 9

Stima dei Parametri Metodo dei Minimi Quadrati

9.1 Metodo dei Minimi Quadrati

In questo paragrafo illustriamo *metodo dei minimi quadrati*, uno dei metodi di stima puntuale più utilizzati in fisica e non solo. Consideriamo un esperimento nel quale siano state acquisite N coppie di valori (x_i, y_i) di due grandezze fisiche X e Y tra le quali si ipotizza che esista una relazione funzionale[1] $Y = f(X|\Lambda)$, (il *modello*). Come già detto la funzione f dipenderà anche da un certo numero di parametri $\Lambda : \{\lambda_k\}$; ad esempio se la relazione funzionale è lineare avremo: $f(X|\lambda_1, \lambda_2) = \lambda_1 + \lambda_2 X$, se è quadratica: $f(X|\lambda_1, \lambda_2, \lambda_3) = \lambda_1 + \lambda_2 X + \lambda_3 X^2$, se è esponenziale: $f(X|\lambda_1, \lambda_2) = \lambda_1 \exp(\lambda_2 X)$ e così via. Supponiamo inoltre le grandezze y_i siano affette dalle incertezze u_{y_i} mentre le incertezze sulle grandezze x_i possano essere ritenute trascurabili.

Fatte queste premesse, possiamo enunciare il **Principio dei Minimi Quadrati**: la stima migliore dei parametri incogniti Λ è quella che *minimizza la seguente espressione*:

$$R^2 = \sum_{i=1}^{N} \frac{(y_i - f(x_i|\Lambda)\,)^2}{u_{y_i}^2} \tag{9.1}$$

somma R^2 degli scarti quadrati tra il valore sperimentale e quello previsto dal modello (il valore “teorico”) diviso per la stima della varianza di y_i ($u_{y_i}^2$). Per ottenere la stima di parametri Λ si deve trovare il minimo dell'espressione (9.1) vista come funzione dei parametri, ovvero si deve risolvere il seguente sistema:

$$\frac{\partial R^2}{\partial \lambda_1} = \frac{\partial R^2}{\partial \lambda_2} = \ldots = \frac{\partial R^2}{\partial \lambda_m} = 0 \tag{9.2}$$

dove m è il numero dei parametri da determinare. Se indichiamo con $\Lambda_o = \{\lambda_{1o}, \lambda_{2o}, \ldots \lambda_{mo}\}$ la soluzione del sistema (9.2), la funzione $f(X, \Lambda_o)$ è completamente determinata e generalmente viene indicata con il termine gergale di *fit*[2] dei dati.

[1] La relazione matematica da utilizzare può derivare da considerazioni teoriche legate all'analisi del fenomeno che si studia oppure può essere suggerita dagli andamenti fenomenologici sperimentali

[2] Il termine inglese *fit* si può tradurre, in questo contesto, con: procedura di adattamento di una curva teorica ai dati sperimentali

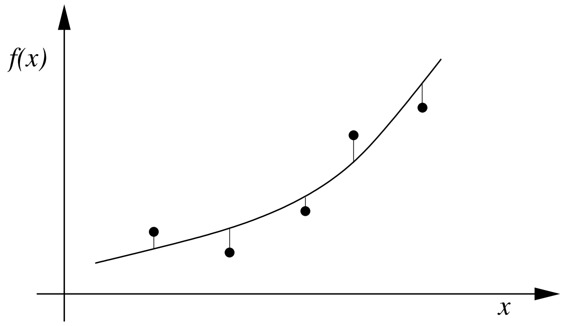

Figura 9.1: La somma delle distanze al quadrato tra i punti "sperimentali" e i corrispondenti valori teorici (quelli sulla curva) deve essere minima per la definizione della *migliore curva* secondo il Metodo del Minimi Quadrati.

Si noti il principio dei minimi quadrati *non richiede alcuna condizione sulla distribuzione di probabilità delle grandezze* y_i; tuttavia se le y_i hanno *distribuzioni gaussiane* con valore medio $f(x_i|\Lambda_o)$ e deviazione standard u_{y_i} allora l'espressione che si ottiene dalla (9.1) quando $\Lambda = \Lambda_o$ è per definizione una variabile che segue la distribuzione del χ^2 (vedi il paragrafo 5.9.6), e potremo scrivere[3]:

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - f(x_i|\Lambda_o)\,)^2}{u_{y_i}^2} \tag{9.3}$$

In questo caso, utilizzando le tabelle precompilate della distribuzione del χ^2, si potranno fare considerazioni probabilistiche sulla qualità del *fit* ottenuto[4], come sarà mostrato nel capitolo 10.

9.2 Minimi quadrati e fit lineare

L'applicazione più semplice del metodo dei minimi quadrati si ha quando la funzione che si vuole adattare all'insieme delle N coppie di dati $(x_i, y_i \pm u_{y_i})$ è lineare. Il modello teorico in questo caso è rappresentato da una retta: $y = a + bx$ (per raccordarsi con la notazione precedente: $\lambda_1 = a$ e $\lambda_2 = b$), nel piano xy e la funzione R da minimizzare si scrive come:

$$R^2 = \sum_{i=1}^{N} w_i (y_i - a - bx_i)^2$$

dove per semplificare la notazione si sono introdotti i pesi w_i delle misure y_i definiti dalla relazione: $w_i = 1/u_{y_i}^2$. In valori di a e b che minimizzano R^2 sono quelli che annullano le derivate di R^2 rispetto ai due parametri:

$$\begin{cases} \dfrac{\partial R^2}{\partial a} = -\,2 \displaystyle\sum_{i=1}^{N} w_i (y_i - a - bx_i) = 0 \\ \dfrac{\partial R^2}{\partial b} = -\,2 \displaystyle\sum_{i=1}^{N} w_i x_i (y_i - a - bx_i) = 0 \end{cases} \tag{9.4}$$

[3]In molti testi, anche se in modo improprio e che può generare confusione, l'espressione (9.1) è indicata con il simbolo χ^2 anche nel caso generale in cui la distribuzione delle y_i non sia normale.

[4]Se la distribuzione cui appartengono le y_i non è gaussiana ma è nota, il calcolo della qualità del fit ottenuto è molto più complessa da ottenere

Per alleggerire la notazione poniamo:

$$S_0 = \sum w_i, \quad S_x = \sum w_i x_i, \quad S_{xx} = \sum w_i x_i^2, \quad S_{xy} = \sum w_i x_i y_i, \quad S_y = \sum w_i y_i \tag{9.5}$$

Con queste posizioni il sistema precedente si scrive:

$$\begin{cases} aS_0 + bS_x = S_y \\ aS_x + bS_{xx} = S_{xy} \end{cases} \tag{9.6}$$

La soluzione del sistema precedente è:

$$\hat{a} = \frac{1}{\Delta}\left(S_{xx}S_y - S_x S_{xy}\right), \qquad \hat{b} = \frac{1}{\Delta}\left(S_0 S_{xy} - S_y S_x\right) \tag{9.7}$$

dove

$$\Delta = S_0 S_{xx} - (S_x)^2$$

Si verifica facilmente che i valori trovati corrispondono al minimo di R^2. Riassumendo, le relazioni (9.7) sono le stime dei parametri a e b ottenute con il metodo dei minimi quadrati.

9.2.1 Incertezza sulle stime dei parametri

Le stime (9.7) dei parametri a e b sono variabili aleatorie in quanto funzioni, in particolare lineari, delle variabili aleatorie y_i e le loro incertezze u_a e u_b si ottengono dalla formula di propagazione delle incertezze[5] (7.10). Inoltre è evidente che $\hat{a}$ e $\hat{b}$ sono grandezze correlate tra loro in quanto entrambe dipendono dalle stesse variabili aleatorie y_i, quindi la loro covarianza, $\mathbb{C}\text{ov}[\hat{a},\hat{b}]$, è diversa da zero. Valuteremo quindi sia l'incertezza delle stime $\hat{a}$ e $\hat{b}$ sia la loro covarianza.

Come è stato detto $\hat{a}$ e $\hat{b}$ sono funzioni lineari delle y_i, quindi potremo scrivere $\hat{a} = \sum_i \alpha_i y_i$ e $\hat{b} = \sum_i \beta_i y_i$, dove le espressioni delle costanti α_i e β_i si ricavano dalle (9.7) e dalle posizioni (9.6): e valgono

$$\begin{aligned} \alpha_i =& \frac{1}{\Delta}\left((\sum_k w_k x_k^2) w_i - (\sum_k w_k x_k) w_i x_i \right) = \frac{w_i}{\Delta}\left(S_{xx} - S_x x_i\right) \\ \beta_i =& \frac{1}{\Delta}\left((\sum_k w_k) w_i x_i - (\sum_k w_k x_k) w_i \right) = \frac{w_i}{\Delta}\left(S_o x_i - S_x\right) \end{aligned}$$

Tenendo conto che $\partial y_k / \partial y_i = \delta_{ki}$, poiché le y_i sono indipendenti, e che per definizione $u_i^2 = 1/w_i$, calcoliamo l'incertezza su a come:

$$\begin{aligned} u_a^2 =& \sum_{i=1}^{N} \left(\frac{\partial a}{\partial y_i}\right)^2 u_{y_j}^2 = \sum_{i=1}^{N} \left(\frac{\partial \sum_k \alpha_k y_k}{\partial y_i}\right)^2 u_{y_i}^2 = \sum_{i=1}^{N} \frac{\alpha_i^2}{w_i} = \sum_{i=1}^{N} \frac{w_i}{\Delta^2}\left(S_{xx} - S_x x_i\right)^2 = \\ =& \frac{1}{\Delta^2} \sum_{i=1}^{N} w_i \left(S_{xx}^2 + S_x^2 x_i^2 - 2 S_{xx} S_x x_i\right) = \frac{1}{\Delta^2}\left(S_o S_{xx}^2 + S_x^2 S_{xx} - 2 S_{xx} S_x^2\right) = \\ =& \frac{S_{xx}}{\Delta} \end{aligned} \tag{9.8}$$

[5] La formula utilizzata è quella per la propagazione delle incertezze non correlate in quanto le y_i sono tra loro indipendenti.

Procedendo allo stesso modo per il calcolo dell'incertezza su b si ottiene

$$
\begin{aligned}
u_b^2 =& \sum_{j=1}^{N}\left(\frac{\partial b}{\partial y_j}\right)^2 u_{y_j}^2 = \sum_{i=1}^{N}\left(\frac{\partial \sum_k \beta_k y_k}{\partial y_i}\right)^2 u_{y_i}^2 = \sum_{i=1}^{N}\frac{\beta_i^2}{w_i} = \sum_{i=1}^{N}\frac{w_i}{\Delta^2}\left(S_o x_i - S_x\right)^2 = \\
=& \frac{1}{\Delta^2}\sum_{i=1}^{N} w_i\left(S_o^2 x_i^2 + S_x^2 - 2S_o S_x x_i\right) = \frac{1}{\Delta^2}\left(S_o^2 S_{xx} + S_o S_x^2 - 2S_o S_x^2\right) \\
=& \frac{S_0}{\Delta}
\end{aligned}
\tag{9.9}
$$

Calcoliamo ora la covarianza $\mathbb{C}\mathrm{ov}[\hat{a}, \hat{b}]$ tra le stime dei parametri $\hat{a}$ e $\hat{b}$:

$$
\mathbb{C}\mathrm{ov}[\hat{a}, \hat{b}] = \mathbb{C}\mathrm{ov}\left[\sum_i \alpha_i y_i, \sum_j \beta_j y_j\right] = \sum_{i,j}\alpha_i\beta_j\,\mathbb{C}\mathrm{ov}[y_i, y_j]
$$

dove nell'ultimo passaggio si è usata una proprietà dell'operatore covarianza che si deduce facilmente usando la definizione di covarianza[6]. Tenendo conto che per l'indipendenza delle y_i si ha $\mathbb{C}\mathrm{ov}[y_i, y_j] = \delta_{ij}\,\mathbb{V}\mathrm{ar}[y_i]$, otteniamo

$$
\begin{aligned}
\mathbb{C}\mathrm{ov}[\hat{a}, \hat{b}] =& \sum_{i,j}\alpha_i\beta_j\,\mathbb{C}\mathrm{ov}[y_i, y_j] = \sum_i \alpha_i\beta_i\,\mathbb{V}\mathrm{ar}[y_i] = \sum_i \frac{\alpha_i\beta_i}{w_i} = \frac{1}{\Delta^2}\sum_i w_i\left(S_{xx} - S_x x_i\right)\left(S_o x_i - S_x\right) = \\
=& \frac{1}{\Delta^2}\sum_i w_i(S_{xx}S_o x_i - S_{xx}S_x - S_o S_x x_i^2 + S_x^2 x_i) = \\
=& -\frac{S_x}{\Delta}
\end{aligned}
\tag{9.10}
$$

Dalla covarianza (9.10) e dalle (9.8) e (9.9) possiamo ottenere il coefficiente di correlazione tra le stime dei parametri $\hat{a}$ e $\hat{b}$

$$
\rho_{ab} = \frac{\mathbb{C}\mathrm{ov}[\hat{a}, \hat{b}]}{\sqrt{\mathbb{V}\mathrm{ar}[\hat{a}]\,\mathbb{V}\mathrm{ar}[\hat{b}]}} = -\frac{S_x}{\Delta}\sqrt{\frac{\Delta}{S_{xx}}\frac{\Delta}{S_o}} = -\frac{S_x}{\sqrt{S_{xx}S_o}}
$$

Riassumendo le incertezze standard e il coefficiente di correlazione dei parametri a e b stimati con il metodo dei minimi quadrati sono:

$$
u_a = \sqrt{\frac{S_{xx}}{\Delta}} \qquad u_b = \sqrt{\frac{S_0}{\Delta}} \qquad \rho_{ab} = -\frac{S_x}{\sqrt{S_{xx}S_o}}
\tag{9.11}
$$

In vari testi di analisi dati che trattano il metodo dei minimi quadrati applicato al fit lineare sono valutate unicamente le incertezze u_a e u_b e non è riportata la covarianza. Tuttavia se $\hat{a}$ e $\hat{b}$ sono usati nella stessa espressione per la valutazione di una grandezza la conoscenza della covarianza è necessaria per la corretta valutazione dell'incertezza della grandezza.

[6] Dimostriamo questa proprietà calcolando esplicitamente la covarianza della variabile aleatoria x_1 moltiplicata per la costante a con una combinazione lineare delle variabili aleatorie y_1 e y_2: $by_1 + cy_2$.

$$
\begin{aligned}
\mathbb{C}\mathrm{ov}[ax, by_1+cy_2] =& \mathbb{E}[(ax-a\mu_x)(by_1+cy_2-b\mu_{y_1}-c\mu_{y_2})] = \mathbb{E}(ax-a\mu_x)(by_1-b\mu_{y_1})+\mathbb{E}[(ax-a\mu_x)(cy_2-c\mu_{y_2})] \\
=& ab\,\mathbb{C}\mathrm{ov}[x_1, y_1] + ac\,\mathbb{C}\mathrm{ov}[x_1, y_2]
\end{aligned}
$$

Caso delle incertezze uguali. Nel caso in cui tutte le incertezze sulle y_i siano uguali ($u_{y_i} = u_y$) le posizioni (9.5) si semplificano, utilizzando il peso $w = 1/u_y^2$, in

$$S_0 = wN, \quad S_x = w\sum x_i, \quad S_{xx} = w\sum x_i^2, \quad S_{xy} = w\sum x_i y_i, \quad S_y = w\sum y_i \tag{9.12}$$

Si verifica facilmente che in questo caso il valore di entrambi i parametri a e b, dati dalle (9.7), *non dipende dal valore dell'incertezza* u_y. Le incertezze su $\hat{a}$ e $\hat{b}$ e il loro coefficiente di correlazione date dalle (9.11) in questo caso divengono:

$$u_a = u_y\sqrt{\frac{\sum x_i^2}{N\sum x_i^2 - (\sum x_i)^2}}, \quad u_b = u_y\sqrt{\frac{N}{N\sum x_i^2 - (\sum x_i)^2}}, \quad \rho_{ab} = -\frac{\sum x_i}{\sqrt{N\sum x_i^2}}$$

9.2.2 Esempi

Consideriamo un esperimento con cui si voglia misurare la costante elastica di una molla elicoidale. L'esperimento si esegue nel seguente modo: la molla viene sospesa e se ne misura l'allungamento rispetto alla sua posizione di riposo in funzione della massa che è agganciata all'estremo libero della molla. Nell'esperimento si misurano, con un'incertezza trascurabile, la massa del peso che si aggancia all'estremo libero della molla (variabile x) e il conseguente allungamento (variabile y) con la sua incertezza u_y. Nella tabella seguente sono riportate 5 coppie di misure delle grandezze x e y.

x (g)	$(y \pm u_y)$ (mm)
50	50 ± 2
100	91 ± 4
150	120 ± 6
200	174 ± 8
250	225 ± 11

La teoria dell'elasticità, in particolare la legge di Hooke ($F = -k\Delta x$), prevede che in condizioni di equilibrio ci sia proporzionalità tra la forza generata dalla molla e il suo allungamento. Quindi il modello matematico da applicare è quello lineare. Volendo stimare con il metodo dei minimi quadrati il valore della costante k della molla assumeremo come variabile "x" (quella con incertezze trascurabili) la massa del peso e come variabile "y" l'allungamento della molla. Nel grafico in figura 9.2 sono riportati i punti della tabella precedente con la massa del peso nell'asse delle ascisse e l'allungamento della molla nell'asse delle ordinate. La curva (una retta) che si vuole adattare ai dati (ovvero eseguire il "*fit*") è:

$$y = a + bx \tag{9.13}$$

dove la costante a tiene conto del fatto che in questo esperimento non sempre si riesce a misurare il vero *allungamento* (rispetto alla condizione di "molla a riposo"); in pratica si misura la coordinata dell'estremo della molla rispetto ad un'origine arbitraria. La costante a, una volta stimata, fornirà appunto il valore della coordinata dell'estremo della molla a riposo. La costante b, attraverso la legge di Hooke, si esprime come $b = k/g$ dove k è la costante della molla cercata e g è l'accelerazione di gravità.

Applichiamo il metodo dei minimi quadrati a questo problema iniziando con il calcolare il valore delle espressioni (9.5). Con semplici calcoli algebrici si ottiene[7]

[7] Per questo tipo di calcoli, in particolare quando il numero di dati è elevato, l'aiuto di programmi di calcolo facilita e velocizza la procedura e rende più difficile fare errori nei calcoli. Un programma particolarmente utile per questi calcoli è il foglio elettronico excel oppure il suo equivalente *open source*

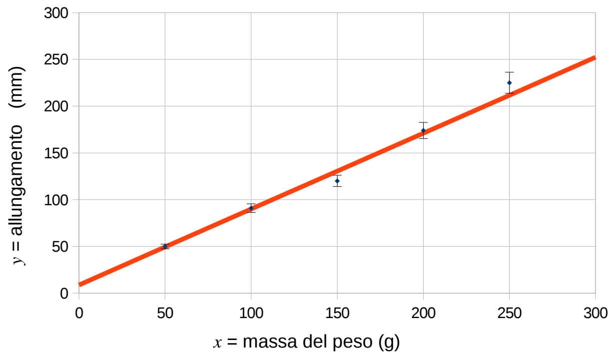

Figura 9.2: Grafico dell'allungamento di una molla in funzione della massa del peso applicato. La retta nel grafico è il *fit* ai dati ottenuto con il metodo dei minimi quadrati.

$$S_0 = 0.257,\ S_x = 21.615,\ S_{xx} = 2530.332,\ S_y = 19.806,\ S_{xy} = 2243.775$$

e quindi

$$\hat{a} = \frac{S_{xx}S_y - S_x S_{xy}}{S_0 S_{xx} - (S_x)^2} = 8.803\,\text{mm} \qquad \hat{b} = \frac{S_0 S_{xy} - S_y S_x}{S_0 S_{xx} - (S_x)^2} = 0.812\,\text{mm g}^{-1} \tag{9.14}$$

Attraverso le (9.11) si ottengono le incertezze standard sui parametri a e b:

$$u_a = \sqrt{\frac{S_{xx}}{S_0 S_{xx} - (S_x)^2}} = 3.71\,\text{mm} \quad u_b = \sqrt{\frac{S_0}{S_0 S_{xx} - (S_x)^2}} = 0.0374\,\text{mm g}^{-1}$$

e il coefficiente di correlazione:

$$\rho_{ab} = -\frac{S_x}{\sqrt{S_{xx}S_o}} = -0.848$$

In conclusione la stima puntuale dei parametri a e b con il metodo dei minimi quadrati è:

$$\hat{a} = (8 \pm 4)\,\text{mm} \qquad \hat{b} = (0.81 \pm 0.04)\,\text{mm g}^{-1} \qquad \rho_{ab} = -0.85$$

Dal valore di b si risale alla costante k della molla attraverso la relazione $k = bg$. L'incertezza su k si ottiene propagando unicamente quella b, solo se si assume per g un valore la cui incertezza sia tale che $u_g \ll g u_b / b$. La giustificazione di questa condizione è lasciata al lettore.

9.2.3 Minimi quadrati e Media pesata

Supponiamo di avere n misure y_i di una stessa grandezza fisica, ciascuna con una propria incertezza u_i. Supponiamo inoltre che queste misurazioni siano state eseguite con diverse modalità per cui le varie incertezze sono differenti, come ad esempio mostrato nella figura 9.3. Applichiamo il metodo dei minimi quadrati per ottenere una stima della grandezza che tenga conto di tutte le misurazioni eseguite con le loro incertezze. In questo caso la funzione $f(x|\Lambda)$ si riduce ad una costante, λ, il cui valore corrisponde al valore della grandezza cercata. La funzione R da minimizzare è

$$R^2 = \sum_{i=1}^{n} \frac{(y_i - \lambda)^2}{u_i^2}$$

derivando e annullando la relazione precedente si ha:

$$\frac{\partial R^2}{\partial \lambda} = -2\sum_{i=1}^{n}\frac{y_i-\lambda}{u_i^2} = -2\left(\sum_{i=1}^{n}\frac{y_i}{u_i^2} - \lambda\sum_{i=1}^{n}\frac{1}{u_i^2}\right) = 0$$

Da cui introducendo i pesi $w_i = 1/u_i^2$ otteniamo la nota formula della *media aritmetica pesata*:

$$\overline{y} \equiv \lambda = \frac{\sum w_i y_i}{\sum w_i}$$

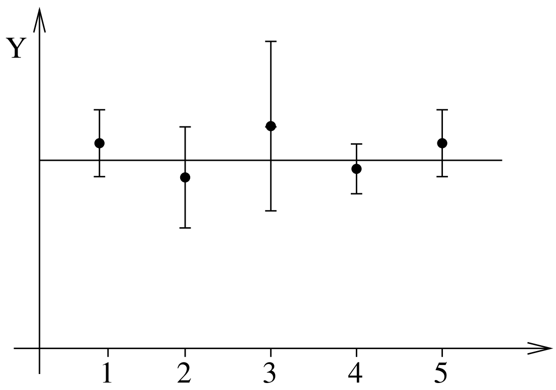

Figura 9.3: Misure ripetute con incertezze differenti.

9.2.4 Minimi quadrati e Massima Verosimiglianza

Il metodo dei minimi quadrati può essere derivato da quello di Massima Verosimiglianza nel caso particolare in cui le grandezze "y" siano distribuite in modo gaussiano. Consideriamo due grandezze X e Y tra le quali si suppone esista una relazione funzionale $Y = \phi(X|\Lambda)$ dove Λ rappresenta l'insieme dei parametri da determinare eseguendo un esperimento. Siano x_i $y_i \pm \sigma_i$ N misurazioni indipendenti delle grandezze X e Y con le y distribuite in modo gaussiano attorno al valore $\phi(x_i|\Lambda)$ con deviazione standard σ_i e siano inoltre trascurabili le incertezze sulle x_i. Per ipotesi, in corrispondenza ad ogni valore x_i le misurazioni y hanno la distribuzione:

$$f_i(y) = \frac{1}{\sqrt{2\pi}\sigma_i}\mathrm{e}^{-(y-\phi(x_i|\Lambda))^2/2\sigma_i^2} \tag{9.15}$$

Per l'indipendenza delle misure y_i la probabilità di ottenere un particolare evento $y_1,\ldots,y_N$ è proporzionale a: $f_1(y_1|\Lambda)f_2(y_2|\Lambda)\ldots f_N(y_N|\Lambda)$. Utilizzando la (8.5), fa funzione di verosimiglianza in questo caso si scrive:

$$\mathcal{L}(\Lambda) = \sum_{i=1}^{N}\ln\frac{1}{\sqrt{2\pi}\sigma_i}\mathrm{e}^{-(y_i-\phi(x_i|\Lambda))^2/2\sigma_i^2} = -\left(\sum_{i=1}^{N}\frac{(y_i-\phi(x_i|\Lambda))^2}{2\sigma_i^2}\right) + cost.$$

Il metodo di massima verosimiglianza richiede di massimizzare $\mathcal{L}$ che, in questo caso, si ottiene minimizzando la sommatoria tra parentesi. Quest'ultima è proprio la somma degli scarti quadrati divisi per la varianza, la cui minimizzazione è esattamente la condizione richiesta dal metodo dei minimi quadrati, a parte l'ininfluente fattore 2. Concludiamo quindi che quando le variabili aleatorie sono distribuite in modo normale, il metodo dei minimi quadrati e quello di massima verosimiglianza sono equivalenti, ovvero portano alla stessa stima dei parametri incogniti.