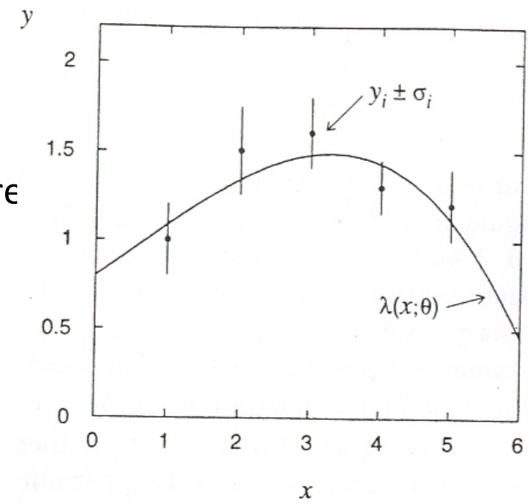
Lezione 7
Metodo dei Minimi Quadrati

Stimatori di Minimi Quadrati

- Supponiamo di misurare due variabili casuali X e Y: ad ogni valore di X misuro il valore di Y. Per esempio negli istanti $x_1, x_2, \dots, x_n$ misuro le posizioni $y_1, y_2, \dots, y_n$. Ognuna di queste misure avrà una propria deviazione standard σ_i



- Supponiamo di conoscere la relazione funzionale $\lambda(x; \theta)$ che per ogni x mi permette di determinare il corrispondente valore di y

- La funzione λ contiene un parametro (o più parametri) che devo determinare a partire dalle misure sperimentali

- Ad ogni misura x_i associamo un valore misurato y_i ed un valore stimato $\lambda(x_i; \theta)$

- La differenza $y_i - \lambda(x_i; \theta)$ è detta residuo. Sommiamo i quadrati dei residui di tutte le misure pesati con l'inverso della loro deviazioni standard σ_i

Stimatori di Minimi Quadrati

- Questa somma è chiamata χ^2

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{y_i - \lambda(x_i;\theta)}{\sigma_i} \right]^2$$

- Il parametro incognito da stimare è il valore che minimizza questa funzione. Questo stimatore è detto dei minimi quadrati (LS)

- Consideriamo il caso che la relazione λ sia di tipo lineare (nei parametri): y = mx con m parametro da stimare. Con n misure della variabile X calcolo il χ^2:

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{y_i - m x_i}{\sigma_i} \right]^2$$

- Per determinare il minimo di questa funzione pongo uguale a zero la derivata prima rispetto al parametro m

$$\frac{\partial \chi^2}{\partial m} = -2 \sum_{i=1}^{n} x_i \frac{y_i - m x_i}{\sigma_i^2} = 0$$

Stimatori di Minimi Quadrati

- Se le misure hanno tutte la stessa varianza σ^2 , allora si ha:

$$-\frac{2}{\sigma^2}\sum_{i=1}^{n}(x_i y_i - m x_i^2) = 0$$

- Il valore del parametro m che annulla questa relazione è dato da :

$$\hat{m} = \frac{\overline{xy}}{\overline{x^2}}$$

- Questo risultato può essere riscritto cosi: $\hat{m} = \sum_{i=1}^{n} \frac{x_i}{n\overline{x^2}} y_i$

- Propagando gli errori da ogni y_i ad m si ha: $V[\hat{m}] = \sum_{i=1}^{n} \left(\frac{x_i}{n\overline{x^2}}\right)^2 \sigma^2 = \frac{\sigma^2}{n\overline{x^2}}$

- Questo risultato si può generalizzare al caso di stimatore di pendenza ed intercetta all'origine : y = ax + b

Massima Verosimiglianza e Minimi Quadrati

- Supponiamo che le distribuzioni delle variabili casuali siano di tipo **gaussiano** e che le n misure y_i siano tra di loro indipendenti

- Indichiamo con $\lambda_i = \lambda(x_i; \theta)$ e y_i il valore stimato e quello misurato di Y corrispondenti alla misura x_i

- La p.d.f. per y_i è quindi: $g(y_i; \lambda_i, \sigma_i^2) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{[y_i - \lambda(x_i;\theta)]^2}{2\sigma_i^2}}$

- Per le n misure la log-likelihood è data da:

$$logL = -\frac{1}{2}\sum_{i=1}^{n} \frac{[y_i - \lambda(x_i;\theta)]^2}{\sigma_i^2} - \sum_{i=1}^{n} log\sigma_i\sqrt{2\pi}$$

- Per massimizzare la log-likelihood bisogna minimizzare la quantità:

$$\chi^2(\theta) = \sum_{i=1}^{n} \frac{[y_i - \lambda(x_i;\theta)]^2}{\sigma_i^2}$$

Massima Verosimiglianza e Minimi Quadrati

- A meno di termini che non contengono i parametri si ha che

$$\chi^2 = - 2\log L$$

  In questo caso lo stimatore di ML e quello dei minimi quadrati forniscono la stessa stima

- Se invece le misure y_i non sono tra di loro indipendenti allora bisogna tener conto dei termini covarianti ed usare la matrice di covarianza V

- Se la matrice V è nota, allora la log-likelihood si scrive così:

$$logL = -\frac{1}{2}\sum_{i,j=1}^{n}(y_i - \lambda(x_i;\theta))V_{ij}^{-1}(y_j - \lambda(x_j;\theta))$$

- Il massimo di questa funzione corrisponde al minimo della funzione

$$\chi^2(\theta) = \sum_{i,j=1}^{n}(y_i - \lambda(x_i;\theta))(V^{-1})_{ij}(y_j - \lambda(x_j;\theta))$$

Proprietà degli Stimatori LS

- A differenza degli stimatori ML, quelli LS non hanno proprietà generali ottimali tranne che nel caso particolare che la relazione funzionale sia di tipo lineare

- Se la relazione funzionale $\lambda(x; \theta)$ è di tipo lineare nei parametri θ allora lo stimatore LS è non distorto

- Questo stimatore è a minima varianza tra tutti gli stimatori che sono funzioni lineari nei parametri

- Questo stimatore viene anche usato quando le singole misure non sono gaussiane. È probabilmente lo stimatore più comunemente usato

- La quantità da minimizzare è detta χ^2 perché sotto determinate condizioni ha una p.d.f. del χ^2. Mantiene questo nome anche quando questo non è vero

Fit Lineari

- Sia λ(x; θ) funzione lineare dei parametri θ = θ(θ_1, θ_2, .., θ_m) da stimare

$$\lambda(x;\theta) = \sum_{j=1}^{m} a_j(x)\theta_j$$

  dove le a_j(x) sono generiche funzioni di x tra di loro linearmente indipendenti

- Sotto queste condizioni i parametri da stimare e le loro varianze si possono trovare analiticamente.

- Possiamo scrivere : $\lambda(x_i;\theta) = \sum_{j=1}^{m} a_j(x_i)\theta_j = \sum_{j=1}^{m} A_{ij}\theta_j$

- col χ^2 che in notazione matriciale si scrive

$$\chi^2 = (\mathbf{y} - \lambda)^T V^{-1} (\mathbf{y} - \lambda)$$

$$= (\mathbf{y} - A\theta)^T V^{-1} (\mathbf{y} - A\theta)$$

Fit Lineari

- I vettori delle misure e dei valori predetti sono vettori colonna

- Per minimizzare il χ2 si annullano le derivate parziali rispetto ai parametri

$$-2(A^T V^{-1}\mathbf{y} - A^T V^{-1} A\theta) = 0$$

- Se la matrice $A^T V^{-1} A$ non è singolare , allora si ha:

$$\hat{\theta} = (A^T V^{-1} A)^{-1} A^T V^{-1}\mathbf{y}$$

che sono i valori dei parametri stimati.

- La matrice di covarianza $U = (A^T V^{-1} A)^{-1}$ si ottiene propagando gli errori delle misure. L'inverso di questa matrice è:

$$(U_{ij})^{-1} = \frac{1}{2}\left[\frac{\partial^2 \chi^2}{\partial\theta_i \partial\theta_j}\right]_{\theta=\hat{\theta}}$$

con le derivate seconde calcolate nei valori stimati dei parametri

Fit Lineari

- Abbiamo già visto che se le misure yi sono di tipo gaussiano vale la relazione $\chi^2 = -2 \log L$. In questo caso la formula vista prima coincide con il limite di Cramer-Rao

- Sempre nella ipotesi di λ lineare nei parametri si può far vedere che il χ^2 è quadratico in θ:

$$\chi^2(\theta) = \chi^2(\hat{\theta}) + \frac{1}{2}\sum_{i,j}^{m}\left[\frac{\partial^2\chi^2}{\partial\theta_i\partial\theta_j}\right]_{\theta=\hat{\theta}}(\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j)$$

- La linee di livello corrispondente al $\chi^2_{min} + 1$ ha tangenti nei punti $\hat{\theta}_i \pm \hat{\sigma}_i$ e fornisce un intervallo di una σ per il parametro stimato

- Se i parametri sono due la linea di livello è una ellisse. Se la funzione λ non è lineare nei parametri, la linea di livello non è ellittica.

LS Fit con Dati Istogrammati

- Supponiamo di avere istogrammato le nostre misure. Siano N il numero di bin dell'istogramma e x_i il valore centrale del bin i-esimo che contiene y_i eventi. n è il numero totale di eventi

- La larghezza dei bin è generalmente la stessa (ma non sempre!)

- Il numero di eventi previsti nel bin i-esimo è dato da

$$\lambda_i(\theta) = n \int_{x_i{}^{min}}^{x_i{}^{max}} \lambda(x;\theta)dx = np_i(\theta)$$

con $p_i(\theta)$ probabilità che l'evento appartenga al bin i-esimo

- I parametri θ li stimiamo minimizzando il χ^2 che scriviamo

$$\chi^2(\theta) = \sum_{i=1}^{N} \frac{(y_i - \lambda_i(\theta))^2}{\sigma_i^2}$$

LS Fit con Dati Istogrammati

- Se y_i è molto più piccolo di n allora la variabile y_i può essere considerata poissoniana. La varianza di y_i è il valore aspettato di eventi nel bin i-esimo e quindi:

$$\chi^2(\theta) = \sum_{i=1}^{N} \frac{(y_i - \lambda_i(\theta))^2}{\lambda_i(\theta)} = \sum_{i=1}^{N} \frac{(y_i - np_i(\theta))^2}{np_i(\theta)}$$

- Ovviamente non si può aumentare a dismisura il numero di bin N dell'istogramma perché se si hanno troppo pochi eventi (circa <5) in un bin lo stimatore sbaglia. Il numero N di bin va ottimizzato

- Come varianza possiamo anche utilizzare direttamente il numero di eventi osservati (al posto di quelli stimati) e scrivere:

$$\chi^2(\theta) = \sum_{i=1}^{N} \frac{(y_i - \lambda_i(\theta))^2}{y_i} = \sum_{i=1}^{N} \frac{(y_i - np_i(\theta))^2}{y_i}$$

- Questo metodo è detto dei Minimi Quadrati Modificato (MLS)

Bontà del Fit col LS

- Se le distribuzioni delle variabili sono gaussiane e per grandi campioni di dati, LS e ML danno gli stessi risultati

- Se inoltre la dipendenza funzionale dell'ipotesi λ è corretta (forma lineare nei parametri) il minimo del χ^2 calcolato segue la distribuzione del $\chi 2$ con $n_d = N - m$ gradi di libertà

- Questo χ^2 può essere usato come test di bontà del fit. Come P-value si considera la probabilità che l'ipotesi fatta abbia un χ^2 uguale o maggiore di quello χ^2_0 trovato nel fit :

$$P = \int_{\chi^2_0}^{\infty} f(z; n_d) dz$$

- Nella distribuzione del χ^2 il valore di aspettazione è uguale a n_d. Allora mi aspetto che χ^2/n_d (detto χ^2 ridotto) sia circa 1

- Se il χ^2 ridotto è circa 1 allora OK. Se non lo è, c'è qualche problema (spesso ciò è dovuto ad errori o sottostimati o sovrastimati)

Combinazione di Più Esperimenti con LS

- Supponiamo che esistano N misure indipendenti della variabile casuale Y, $y_i \pm \sigma_i$
- Sia λ il valore vero aspettato. Allora si ha:

$$\chi^2(\lambda) = \sum_{i=1}^{N} \frac{(y_i - \lambda)^2}{\sigma_i^2}$$

- Azzerando la derivata rispetto a λ e risolvendo per λ, si ha:

$$\lambda = \frac{\sum_{i=1}^{N} y_i/\sigma_i^2}{\sum_{j=1}^{N} 1/\sigma_j^2}$$

cioè la media combinata si ottiene pesando le misure con le varianze

- Passando alle derivate seconde si ha la varianza del valore combinato:

$$V[\hat{\lambda}] = \frac{1}{\sum_{i=1}^{N} 1/\sigma_i^2}$$

- Questa procedura può essere generalizzata a variabili correlate tra di loro tenendo conto della matrice di covarianza