

Ancora su Likelihood/ χ^2 . Likelihood Fit di grafici. Descrizione non parametrica di dati

Laboratorio di Metodi Computazionali e Statistici (2023/2024)

R. Cardinale, F. Parodi, S. Passaggio

December 10, 2023

Legame χ^2 e $\log L$ (y_i, μ_i, σ_i)

Per Gaussiane

$$\mathcal{L} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(y_i - \mu_i)^2}{2\sigma_i^2}}$$

$$\begin{aligned}-\ln \mathcal{L} &= -\sum_i \ln \left(\frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(y_i - \mu_i)^2}{2\sigma_i^2}} \right) = -\sum_i \left[\ln \frac{1}{\sqrt{2\pi}\sigma_i} - \frac{(y_i - \mu_i)^2}{2\sigma_i^2} \right] = \\&= \frac{1}{2} \sum_i \frac{(y_i - \mu_i)^2}{\sigma_i^2} + \text{cost} = \frac{\chi^2}{2} + \text{cost}\end{aligned}$$

$$\Delta \chi^2 = 2 \Delta (-\ln \mathcal{L})$$

Legame χ^2 e $\log L$

$\vec{\mu}$ valori veri $\vec{\mu}'$ valori "alternativi"

$$\chi^2 = \sum \frac{(\mu'_i - \mu_i)^2}{\sigma_i^2} \quad \text{quanto } \mu' \text{ compatibili con } \mu$$

con \mathcal{L} ?

$$-\ln \mathcal{L}(\vec{\mu}) = -\sum_{i=1}^n \left[\ln \frac{1}{\sqrt{2\pi}\sigma_i} - \frac{(\mu_i - \mu_i)^2}{2\sigma_i^2} \right] = -\sum \left[\ln \frac{1}{\sqrt{2\pi}\sigma_i} \right]$$

$$-\ln \mathcal{L}(\vec{\mu}') = -\sum_{i=1}^n \left[\ln \frac{1}{\sqrt{2\pi}\sigma_i} - \frac{(\mu'_i - \mu_i)^2}{2\sigma_i^2} \right]$$

sommo i.e. 2

$$-\ln \mathcal{L}(\vec{\mu}') + \ln \mathcal{L}(\vec{\mu}) = \frac{1}{2} \chi^2(\vec{\mu}')_{\mu} \quad \mu' \text{ rispetto } \mu$$

$$-(\ln \mathcal{L}(\vec{\mu}') - \ln \mathcal{L}(\vec{\mu})) = -\ln \frac{\mathcal{L}(\vec{\mu}')}{\mathcal{L}(\vec{\mu})} = \frac{1}{2} \chi^2(\vec{\mu}')_{\mu}$$

Fit di grafici con errore su x: errore efficace

Vogliamo applicare il metodo di Likelihood fit a grafici con errore non trascurabile sulle x e dedurre la formula dell'errore efficace.

Scriviamo la likelihood assumendo distribuzioni gaussiane (non correlate) sia per la variabile x che per la variabile y . Chiamo le misure $x_{0,i}$, $y_{0,i}$, $\sigma_{x,i}$, $\sigma_{y,i}$.

$$\mathcal{L}(\beta, x) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_{y,i}} \exp \left[-\left(\frac{y_{0,i} - f(x_i; \beta)}{2\sigma_{y,i}^2} \right)^2 \right] \frac{1}{\sqrt{2\pi}\sigma_{x,i}} \exp \left[-\left(\frac{x_{0,i} - x_i}{2\sigma_{x,i}^2} \right)^2 \right]$$
$$\chi^2(\beta, x) = -2\ln\mathcal{L} = \sum_{i=1}^N \left[\left(\frac{y_{0,i} - f(x_i; \beta)}{\sigma_{y,i}} \right)^2 + \left(\frac{x_{0,i} - x_i}{\sigma_{x,i}} \right)^2 \right]$$

Se errore non trascurabile ho prod tra gaussiane. Ora ho $f(x_i, \beta)$ non $f(x_{0,i}, \beta)$

Quindi oltre ai parametri β dovremmo minimizzare anche rispetto ai valori "veri" di x_i . Questo è possibile tramite costruzione esplicita del χ^2 ed il metodo, teoricamente, più preciso.

Tuttavia è anche possibile integrare la Likelihood sull'variabili x_i (in modo da inglobarne l'effetto senza essere obbligati a calcolarne il valore) ottenendo:

$$\mathcal{L}(\beta) = \int \mathcal{L}(\beta, x) dx_1 dx_2 \dots dx_n$$

uso stime x_i solo
per grafico dopo fit

Fit di grafici con errore su x: errore efficace

Motivazione Euristiche su utilizzo errore sull'x

$$\chi^2 = \sum_{i=1}^N \frac{(y_{o,i} - f(x_{o,i}; \beta))^2}{\sigma_i^2} \quad \begin{array}{l} y_o \text{ } x_o \text{ misurato} \\ y \text{ no errore} \end{array}$$

$\sigma_i^2 \rightarrow$ varianza della misura e varianza del misuratore

$$x_{o,i} \pm \sigma_{x_i} \quad \sigma_{y_i}^2 \sim \sigma_{x_i}^2 + \left(\frac{\partial f}{\partial x_i} \Big|_{x_{o,i}} \right)^2 \sigma_{x_i}^2$$

Applicazione lineare funz intorno x_i :

$$f(x_i) = f(x_o) + \frac{\partial f}{\partial x_i} \Big|_{x_i=x_{o,i}} (x_i - x_{o,i}) + o((x_i - x_{o,i})^2) \quad \text{pongo } \delta x_i = x_i - x_{o,i}$$

$$\mathcal{L} \propto \prod_{i=1}^N \exp \left[-\frac{(y_{o,i} - f(x_i))^2}{2\sigma_{y_i}^2} \right] \exp \left[-\frac{(\delta x_i)^2}{2\sigma_{x_i}^2} \right] dx_1 \dots dx_n$$

$$\mathcal{L} \propto \int \exp \left[-\frac{y_o - f(x_o) - f'(x_o) \delta x}{2\sigma_y^2} \right] \exp \left[-\frac{(\delta x)^2}{2\sigma_x^2} \right] d\delta x \quad \begin{array}{l} \text{tutto i } x_k \text{ poi} \\ \text{tutti uguali come} \\ \text{integrali} \end{array}$$

$$\int e^{-\frac{(a-bx)}{2\sigma_x^2}} e^{-\frac{x^2}{2\sigma_x^2}} d\delta x = e^{-\frac{a^2}{2(b^2\sigma_x^2 + \sigma_x^2)}}$$

completare quadrati:

Fit di grafici con errore su x: errore efficace

$$\mathcal{L} \propto \exp \left[\frac{(y_0 - f(x_0))^2}{2[\sigma_y^2 + (f'(x_0))^2 \sigma_x^2]} \right]$$

$$\chi^2 = -2 \ln \mathcal{L} + \text{cost}$$

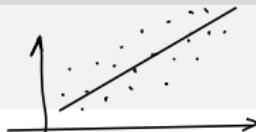
$$\chi^2(\beta) = \sum_{i=1}^n \frac{(y_i - f(x_{0,i}; \beta))^2}{\sigma_{y_i}^2 + (f'(x_i)|_{x_{0,i}})^2 \sigma_{x_i}^2}$$

Likelihood con approx buttando na x mi rida eq dei fit semplici
(metodo ↓ errore efficace) che si usa tutti i giorni
deriva dal metodo generale Likelihood

Fit di grafici con errore su x: errore efficace

TGraphError fit fa media fra errore low e high xk nelle casi
di errori assimmetrici

Fit di grafici senza errore



Come impostare il fit di un grafico senza errori ?

mettre cov olzonli

Si assume che gli errori (ignoti) siano tutti uguali e si pone inizialmente $V = I$.

Per cui:

$$\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}$$

$$V_{\hat{\beta}} = (X^T X)^{-1}$$

Il minimo di χ^2 indica lo scostamento dei vari punti sperimentali dalla curva fittata. L' errore costante può essere dedotto da:

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - f(x_i; \beta))^2}{\sigma_y^2}$$

$$\sigma_{\text{res}}^2 = \frac{\chi^2_{\text{min}}}{n} = \frac{\chi^2}{n}$$

chi² ridotto è la media degli scarti in quadratura
spread dei dati usato come incertezza

La nuova matrice di covarianza è:

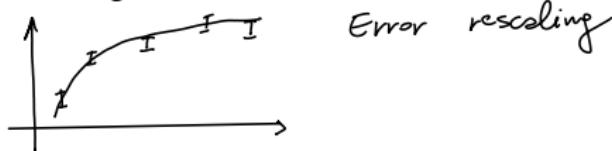
$$V = \sigma_{\text{res}}^2 I$$

Il valore centrali fissati rimangono invariati (cambia il loro errore), il χ^2 ridotto del fit finale è, per costruzione, uguale a 1.

$$\frac{\chi'^2}{n} = \frac{1}{n} \sum_{i=1}^n \frac{(y_i - f(x_i; \beta))^2}{\sigma_{\text{res}}^2}$$

Fit con χ^2 troppo grande

Poco consigliato come metodo



- Un fit con un χ^2 troppo grande, anche se la funzione di fit sembra descrivere correttamente i dati, potrebbe essere sintomo di errori sottostimati.
- Uno dei metodi per tenere conto di questa sottostima e ottenere stime ragionevoli per gli errori dei parametri del fit è quello di riscalare gli errori dei dati in input in modo che il χ^2 ridotto valga 1.
- Chiaramente è preferibile, in generale, se possibile, trovare le ragioni della sottostima dell'incertezza.

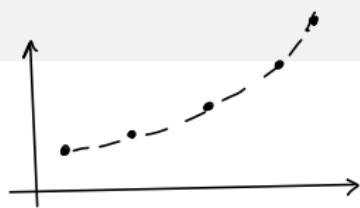
Usato in alcune misure G

Interpolazione e smoothing

- Fit: descrizione dei dati tramite un modello \rightarrow estrae più info dai dati
- Interpolazione: determinare il valore aspettato in una regione in cui non ci sono dati (estrapolazione come interpolazione ma in regioni “esterne”: molto più arbitrario) \hookrightarrow più difficile
- Smoothing: descrizione dei dati senza associare un particolare modello.

Polinomi di Lagrange

Se dati variano continuamente tra loro uscì polinomi



Per $n + 1$ punti (x_i, y_i) , con

$$i = 0, 1, \dots, n \quad x_i \neq x_j$$

c'è un'unica interpolazione di grado n con

$$p(x_i) = y_i \quad i = 0, 1, \dots, n$$

che si costruisce con i polinomi di Lagrange definiti come:

$$L_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^n \frac{(x - x_j)}{(x_i - x_j)}$$

$$p(x) = \sum_{i=0}^n y_i L_i(x)$$

$$L_1 = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} \quad L_1(x_1) = 1$$

Splines

Una "spline" è un polinomio interpolante una serie di punti con i seguenti vincoli:

- $S(x) = S_i(x)$ per $x_i \leq x \leq x_{i+1}$ con $i = 0, 1, \dots, n - 1$
- $S(x_i) = y_i$ con $i = 0, 1, \dots, n$
- $S_i(x_{i+1}) = S_{i+1}(x_{i+1})$ con $k = 0, 1, \dots, n - 2$ *dove essere continue*

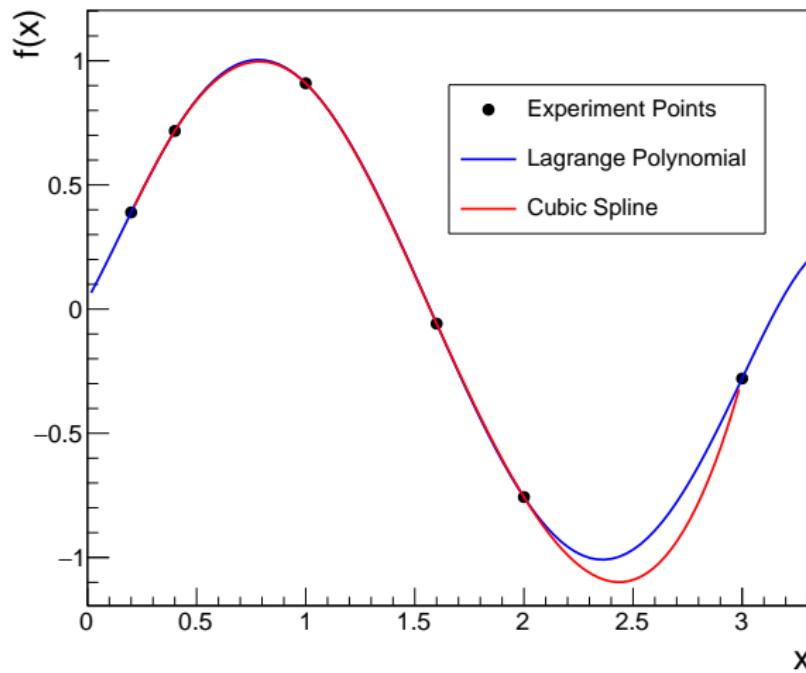
Spline lineare:

$$S_i(x) = y_i + \frac{y_{i+1} - y_i}{x_{i+1} - x_i}(x - x_i)$$

si usa
quadratiche o cubiche

Confronto polinomi di Lagrange/Splines

se campionario ok (Th. Campionamento)



Qui negli zgr.
In 2,5
ma LAGR scende
troppo presto

Lagr fallisce
nei punti distanti
ancoraggi

Smoothing

Scopo dello smoothing è quello fornire una descrizione “compatta” di una serie di punti (x, y) (si applica ugualmente a casi N-dimensional).

KNN (k-Nearest Neighbor) Method

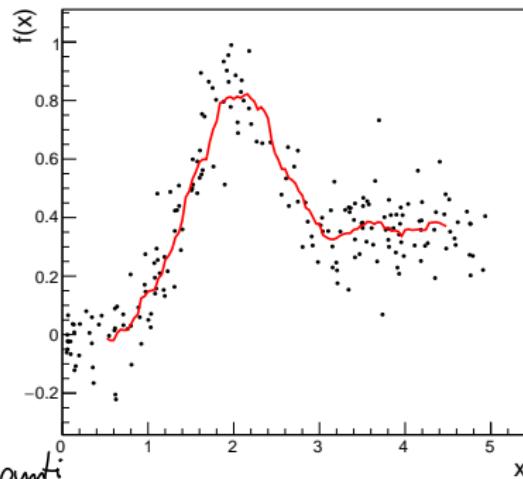
Definisce il valore della funzione come:

$$f(x) = \frac{1}{k} \sum_{i \in N_k(x)} y_i$$

che è la media su k valori in un intorno del punto da stimare.

Tutti i punti hanno lo stesso peso e la funzione è discontinua.

In un intervallo Δx nella media dei punti n° punti cambia tra intervalli \rightarrow è discontinua anche se prende sempre lo stesso valore \uparrow



Kernel smoothing

In alternativa a KNN, si possono assegnare pesi che diminuiscono in funzione della distanza tra x_0 e punti x che contribuiscono alla media usando una kernel funzione:

$$\leftarrow K_\lambda(x_0, x) = D \left(\frac{|x - x_0|}{\lambda} \right)$$

Prendo medi sulle
punti pesata con la
distanza del punto

il parametro λ (detta anche parametro di larghezza di banda) definisce la in cui larghezza efficace della finestra. \rightarrow va provato, non so a priori valuto tutto \rightarrow o leggo f cost, ma colgo ardentemente

- λ grande medo tutto \rightarrow o leggo f cost, ma colgo ardentemente
- λ piccolo prendo 1 punto \rightarrow o leggo stessa f di partenza

Oppure non
smussa
la funzione

La media kernel-pesata di Nadaraya-Watson è definita come:

$$f(x_0) = \frac{\sum_{i=1}^n K_\lambda(x_0, x_i) y_i}{\sum_{i=1}^n K_\lambda(x_0, x_i)}$$

che fornisce una funzione smooth (se lo è quella del kernel).

Come funzione di kernel si usa spesso una gaussiana (in figura con $\sigma = 0.2$).

$$K_\lambda = \frac{1}{\sqrt{2\pi}\lambda} \exp\left(-\frac{(x-x_0)^2}{2\lambda^2}\right)$$

