呂浩宇

Introduction to Data Analysis 2023. Mar. 03

D. Nonlinear problem (用scipy實作,見課程網頁之jupyter notebook)

Problems that depend nonlinearly on the parameters a_k.
For example, fitting 1D Gaussian distribution

(i) when our initial guess of a_k is close to the minimum of χ^2

We can use Taylor expansion to approximate the problem by a quadratic form

$$\chi^2(\vec{a}) \approx \gamma - \vec{d}\cdot\vec{a} + \frac{1}{2}\vec{a}\cdot\overleftrightarrow{D}\cdot\vec{a}$$

for an initial guess $\vec{a}_{cur}$, the step to jump to the best-fit is given by

$$\vec{a}_{min} = \vec{a}_{cur} + \overleftrightarrow{D}^{-1}\cdot\left[-\underbrace{\nabla\chi^2(\vec{a}_{cur})}_{\text{固定}y_i\text{,視}\chi^2\text{為}a_k\text{的函數}}\right]$$

defining $\delta\vec{a} = \vec{a}_{min} - \vec{a}_{cur} \Rightarrow \sum_{l=0}^{M-1}\alpha_{kl}\,\delta a_l = \beta_k$ ——— Eq.(2)

e.g. 一維問題:

$$\chi^2(a) = \gamma - da + \frac{1}{2}Da^2$$

$$\frac{d\chi^2}{da} = Da - d \Rightarrow d = Da_{cur} - \left.\frac{\partial\chi^2}{\partial a}\right|_{a_{cur}}$$

at the minimum of χ^2, $\left.\frac{d\chi^2}{da}\right|_{a_{min}} = 0$

$$\Rightarrow Da_{min} - d$$

$$= Da_{min} - Da_{cur} + \left.\frac{\partial\chi^2}{\partial a}\right|_{a_{cur}} = 0$$

$$\Rightarrow a_{min} = a_{cur} + \frac{-1}{D}\left.\frac{\partial\chi^2}{\partial a}\right|_{a_{cur}}$$

任意維度之問題求 a_{min} 之方法與概念皆無異一維問題
(投影到 χ^2 的 gradient 方向之後即成為一維問題)

Defining

$$\beta_k \equiv -\frac{1}{2}\frac{\partial\chi^2}{\partial a_k}$$

$$\alpha_{kl} \equiv \frac{1}{2}\frac{\partial^2\chi^2}{\partial a_k\partial a_l}$$

$$\Rightarrow \overleftrightarrow{\alpha} = \frac{1}{2}\overleftrightarrow{D}$$

covariant matrix

$$\overleftrightarrow{C} \equiv \overleftrightarrow{\alpha}^{-1}$$

由於近 χ^2 極小處無異線性問題,故 a_k 的variance 仍由 $\overleftrightarrow{C}$ 的對角項給出 (在 χ^2 極小處取 $\overleftrightarrow{C}$ 值)

(ii) when we are far from the minimum of χ^2

step down using the steepest descent method

$$\vec{a}_{next} = \vec{a}_{cur} - \underbrace{constant}_{\text{步幅}} \times \left.\nabla\chi^2\right|_{a_{cur}}$$

defining $\delta\vec{a} = \vec{a}_{min} - \vec{a}_{cur} \Rightarrow \delta a_l = constant \cdot \beta_l$ ——— Eq.(3)

Levenberg–Marquardt Method

$$\sum_{l=0}^{M-1}\alpha'_{kl}\,\delta a_l = \beta_k, \quad \begin{cases}\alpha'_{jj} = \alpha_{jj}(1+\lambda) \\ \alpha'_{jk} = \alpha_{jk} \quad (j\neq k)\end{cases}$$ ——— Eq.(4)

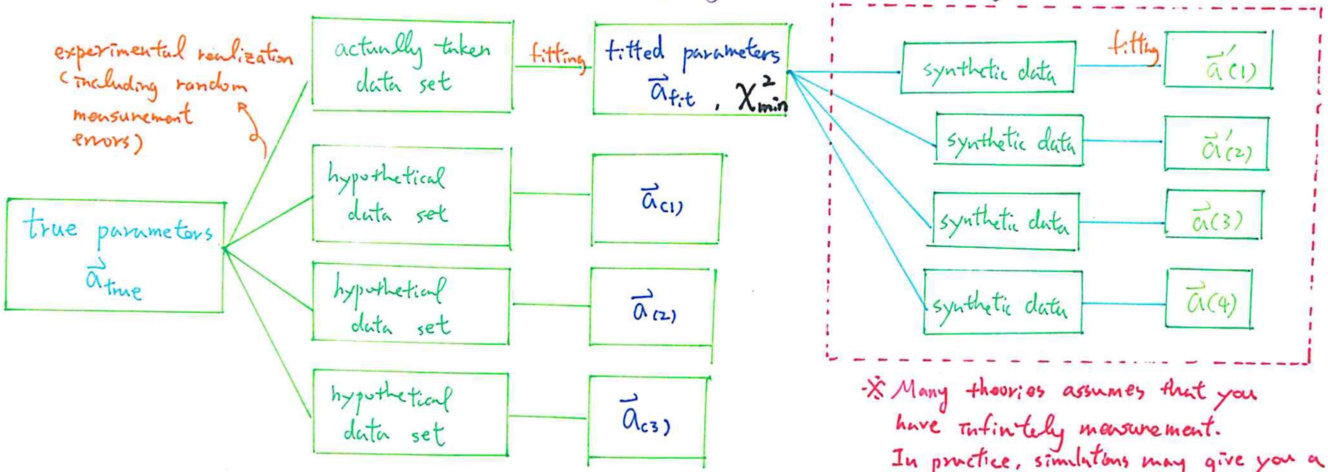
{ when λ is large, diagonal terms in α' dominate, such that Eq.(4) behave like Eq.(3)
{ when λ is small, Eq.(4) behave like Eq.(2)

flow: 選取初始 $\vec{a}$ 及 λ. 若 $\chi^2(\vec{a}+\delta\vec{a}) \geq \chi^2(\vec{a})$, 增加 λ 10倍, 繼續計算 $\delta\vec{a}$, 若 $\chi^2(\vec{a}+\delta\vec{a}) < \chi^2(\vec{a})$, 減小 λ 10倍, 繼續計算 $\delta\vec{a}$

吕浩宇

Introduction to Data Analysis 2023. Mar. 03

E. Practical method for assessing goodness-of-fitting: Bootstraping



※ Many theories assumes that you have infinitely measurement. In practice, simulations may give you a more realistic distribution of the uncertainties

Question: How far $\vec{a}_{fit}$ can be offset from $\vec{a}_{true}$, due to { sparse sampling / thermal noise } (in a statistical sense) ?

Approach: Assume that the errors' behavior around $\vec{a}_{fit}$ is not far from the errors' behavior around $\vec{a}_{true}$. We can then produce synthetic data sets (bootstraping) assuming that the fitted $\vec{a}_{fit}$ is true, and then fit the synthetic data sets to obtain the probability distributions of χ^2 and $\delta\vec{a}$.

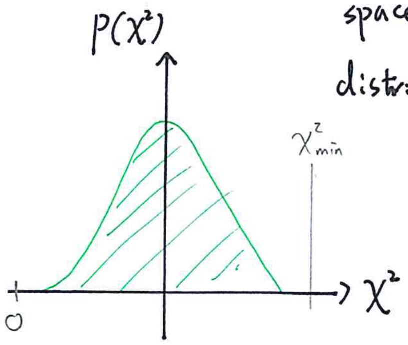
the offset between the fitted parameters and $\vec{a}_{fit}$.

When the errors of the measurements y_i is not Gaussian, this is a practical way of characterizing the uncertainties of a_k.

We can also compare the χ^2_{min} given by $\vec{a}_{fit}$ with the probability distribution of χ^2 obtain from bootstraping.

We can base on contours of iso-χ^2 to define confidence level, e.g. asking what is the chance for the measured χ^2 to be larger than certain values. A confidence region is a region of a M-dim ($\vec{a}$ 的维度) space that contains a certain percentage of the total probability distribution, e.g. 68.3%, 90%, 95.4%, ...

the lowest confidence worthy of quoting



If your χ^2_{min} is located well outside of the 99% confidence region, you may wonder if your model is inappropriate (e.g. the figure on the le

Otherwise, if your χ^2_{min} is within the 10% confidence region, you may wonder whether or not the measurement errors were under-estimated.

呂浩宇

Introduction to Data Analysis   2023. Mar. 03

F. What is the problem with the approach we have been introducing?

1. When you get an upper/lower limit (or many of them, even mostly ...), how do you treat them?
2. when your model is not analytic and the gradient of χ^2 is not defined.?
3. The frequentist inference is incomplete afterall..

→ conceptually, these issues can be better handled with Bayesian inference

↓

A practical algorithm for this purpose is Markov-Chain Monte Carlo method.