Feb 14. 2006

$N > M$

⟨Modeling of Data⟩

N data points (x_i, y_i) $i=1,\cdots,N$

M adjustable parameters a_j $j=1,\cdots,M$.

A "proper" fitting procedure should provide

(i) parameter value

(ii) error estimates on parameter value

(iii) Statistical measures of χ^2 goodness-of-fit → whether the model is appropriate or not.

Common procedure for fitting: least-squares fit.

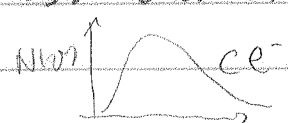
- probability:

  Knowing parameters → prediction of outcome

  n, m, T, v $\qquad \frac{N(v, v+dv)}{V} = \left(\frac{m}{2\pi k_B T}\right)^{3/2} e^{-\frac{1}{k_B T}\left(\frac{1}{2}mv^2\right)} v^2 dv \times 4\pi n$

- likelihood:

  Observation of data → Estimation of parameters



  $c e^{-\frac{1}{k_B T}\left(\frac{1}{2}mv^2\right)}$ $\qquad$ that describe the observed data.

  T, m

Aim of the $\boxed{\text{maximum likelihood estimation}}$ (MLE):

To find the parameter value that makes the observed data most likely.

This is because the likelihood of the parameters given data is defined to be equal to the probability of the data given the parameters.

Suppose that <u>each</u> data point y_i has a measurement error that is independently random and it (y_i) is distributed as a normal distribution (Gaussian function) around the true model ft $y(x_i; a_1, \cdots, a_M)$.

Suppose that the standard deviations σ of these normal distributions are the same for all data points.

Then the probability of obtaining the data set given the parameter values is

$$P = \prod_{i=1}^{N} \left\{ \exp\left[-\frac{1}{2} \left(\frac{y_i - y(x_i)}{\sigma} \right)^2 \right] \right\}$$

maximizing P $\Leftrightarrow$ maximizing $\log P = \sum_{i=1}^{N} \left[-\frac{(y_i - y(x_i))^2}{2\sigma^2} \right]$

||
(Maximum likelihood estimation)

$\Leftrightarrow$ minimizing

$$\sum_{i=1}^{N} \frac{(y_i - y(x_i; a_1, \ldots a_M))^2}{2\sigma^2}$$

M eqs
M unknowns

$\frac{\partial}{\partial a_i}(\quad) = 0$
$i = 1, \cdots M$

Finding the parameter values that describe the given data set the best.

$\Rightarrow$ Least-squares fit

= Finding the prob. of this data set occurring given the parameter values

Use methods for solving linear algebraic eq.
Gauss, LU, SVD

some concents useful for MCS
Mon-3

Review: Basics of a Probability theory

- Random variables : a numerical value that can be mapped into a random event :
  - discrete random variables
    (e.g. # of heads obtained by tossing two coins)
  - continuous random variables
    (e.g. scattering angle of a photon by an atom, or direction of a spin vector in the XY model)

- (Probability) density function : $f(x)$

  A continuous random variable x has a probability density function $f(x)$ if it satisfies the following conditions :

  (i) $f(x)$: single-valued non-negative real number for $x \in \mathbb{R}$

  (ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

  (iii) Probability with which x falls between a and b

  $$: P[a \le x \le b] = \int_a^b f(x)\,dx$$

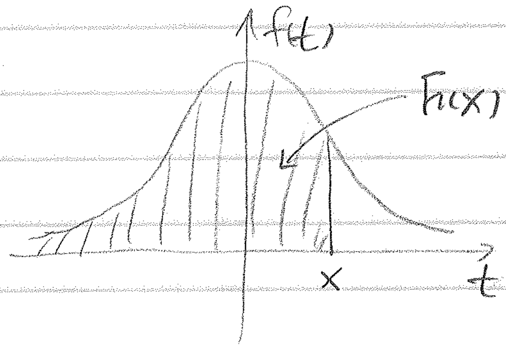
$\Rightarrow f(x)\,dx$: probability that a randomly chosen member from a distribution has a value $(x, x+dx)$

[When one randomly chooses or samples a value from a distribution, $f(x)$ is the prob. that the value is x.]

- (Cumulative) distribution function : $F(x)$

$$F(x) = \int_{-\infty}^{x} dt\, f(t)$$

: probability that a randomly chosen member from a distribution has a value less than x.

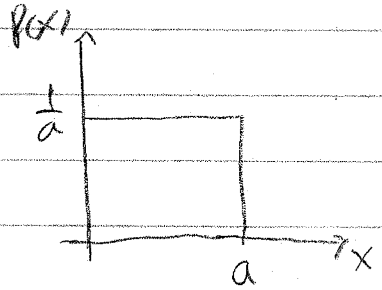
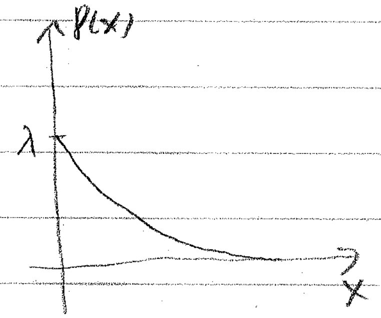
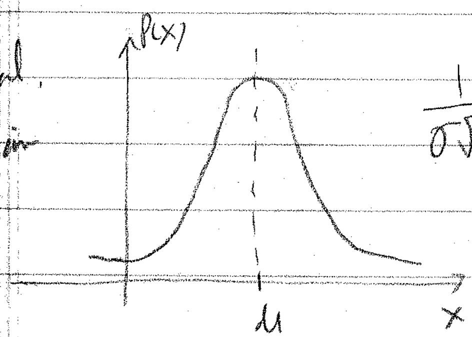


- Examples of Continuous (probability) density function
  - (i) uniform
  - (ii) exponential (Boltzmann)
  - (iii) Normal
  - (iv) χ^2
  - (v) Lorentzian

- Mean : $$\langle x \rangle = \int_{-\infty}^{\infty} dx\, x\, f(x)$$

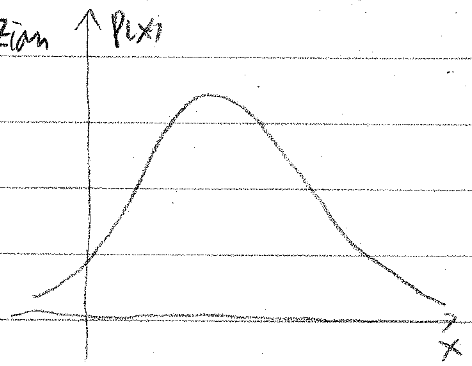
- Variance $$\langle (x - \langle x \rangle)^2 \rangle = \int_{-\infty}^{\infty} dx\, (x - \langle x \rangle)^2 f(x) = \langle x^2 \rangle - \langle x \rangle^2$$

Continuous p.d.f.

		$\langle x \rangle$	$\mathrm{Var}(x)$
(A) uniform $\begin{cases} 0, & x<0,\ x>a \\ \frac{1}{a}, & 0<x<a \end{cases}$		$\frac{1}{2}a$	$\frac{1}{12}a$
(B) exponential (Boltzmann) $\begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x<0 \end{cases}$, $\lambda>0$		$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
(C) Normal, Gaussian $\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$		μ	σ^2

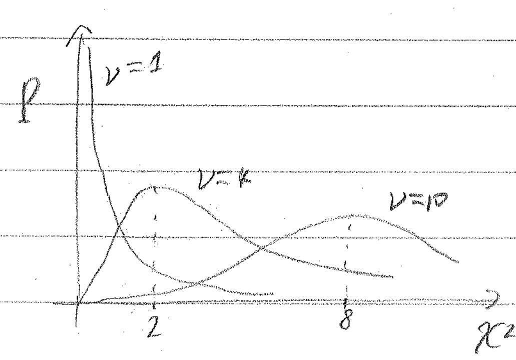
(D) Lorentzian

$$\frac{a}{a^2+x^2}$$



(E) χ^2 : $P(\chi^2, \nu)$

$$\frac{1}{2^{\nu/2}\,\Gamma(\nu/2)} \chi^{2\left[\frac{\nu}{2}-1\right]} e^{-\chi^2/2}, \quad \chi^2>0$$



- Covariance :

x, y : random variable

$\mathrm{cov}\{x,y\} = \langle xy \rangle - \langle x \rangle \langle y \rangle$

If x, y are independent, $\mathrm{cov}\{x,y\} = 0$.

But zero covariance does not guarantee independence of two random variables.

e.g. x, $y = x^2$ $\quad$ $x = \{-1, 0, 1\}$ uniform p.d.f

$\langle x \rangle = 0$

$\langle xy \rangle = \langle x^3 \rangle = 0$.

So $\mathrm{cov}\{x,y\} = \langle xy \rangle - \langle x \rangle \langle y \rangle = 0$ even if

x and y are <u>not</u> independent.

- $\mathrm{var}\{\lambda_1 x + \lambda_2 y\} = \langle \{(\lambda_1 x + \lambda_2 y) - \langle \lambda_1 x + \lambda_2 y \rangle\}^2 \rangle$

$= \langle (\lambda_1 x - \lambda_1 \langle x \rangle)^2 \rangle + \langle (\lambda_2 y - \lambda_2 \langle y \rangle)^2 \rangle$ $\quad$ λ_1, λ_2 : const.

$+ 2 \langle (\lambda_1 x - \lambda_1 \langle x \rangle)(\lambda_2 y - \lambda_2 \langle y \rangle) \rangle$

$= \lambda_1^2 \mathrm{var}(x) + \lambda_2 \mathrm{var}(y) + 2 \lambda_1 \lambda_2 (\langle xy \rangle - \langle x \rangle \langle y \rangle)$

$= \lambda_1^2 \mathrm{var}(x) + \lambda_2 \mathrm{var}(y) + 2 \lambda_1 \lambda_2 \mathrm{cov}\{x,y\}$

If x and y are independent, $\mathrm{cov}\{x,y\} = 0$.

So $\mathrm{var}\{\lambda_1 x + \lambda_2 y\} = \lambda_1^2 \mathrm{var}(x) + \lambda_2^2 \mathrm{var}(y)$

- Correlation coefficient : $\rho(x,y) = \dfrac{\mathrm{cov}\{x,y\}}{\sqrt{\mathrm{var}\{x\} \cdot \mathrm{var}\{y\}}}$, $\quad$ $-1 \le \rho \le 1$

[x, y. indep. $\Rightarrow \rho = 0$]

χ^2 fitting = weighted least-squares fit

chi-square

measurement errors are normally and independently distributed

If each data point (x_i, y_i) has its own standard deviation σ_i, then the maximum likelihood estimate of the model parameters is obtained by minimizing

$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{y_i - y(x_i ; a_1, \cdots, a_M)}{\sigma_i} \right)^2$$

Minimization of χ^2 with respect to $\{a_1, \cdots, a_M\}$ $\Rightarrow$ determine $\{a_1, \cdots, a_M\}$.

Then χ^2 (with determined $\{a_1, \cdots, a_M\}$) is distributed with (probability) density function :

$$f(\chi^2 ; \nu) = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} \chi^{2[(\nu/2)-1]} \exp\left(-\frac{\chi^2}{2}\right) , \quad \chi^2 > 0$$

$\nu = N - M$ # of degrees of freedom

(Cumulative) Distribution ft :

$$P(\chi^2 ; \nu) = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} \int_0^{\chi^2} t^{\frac{\nu}{2}-1} e^{-\frac{t}{2}} \, dt$$

: probability that the chi-square for a set of measurement errors obtained by randomly sampled N observations from Normal distributions is less than χ^2 obtained by fitting our N data points to $\{y(x_i ; a_1, \cdots, a_M)\}$.

Here $\{a_1, \cdots, a_M\}$ are fixed such that χ^2 for our N data points becomes minimum.

$$\frac{\partial \chi^2}{\partial a_k} = 0 \quad (k = 1, \cdots, M) \Rightarrow 0 = \sum_{i=1}^{N} \left(\frac{y_i - y(x_i)}{\sigma_i^2} \right) \left(\frac{\partial y(x_i ; a_1, \cdots, a_M)}{\partial a_k} \right)$$

M equations

M unknowns

$$Q(\chi^2;\nu) = 1 - P(\chi^2;\nu)$$

$$= \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)} \int_{\chi^2}^{\infty} t^{\frac{\nu}{2}-1} e^{-\frac{t}{2}}\, dt$$

$$\frac{t}{2} = u, \quad dt = 2du$$

$$Q(\chi^2;\nu) = \frac{2 \cdot 2^{\frac{\nu}{2}-1}}{2^{\nu/2}\,\Gamma(\nu/2)} \int_{\frac{\chi^2}{2}}^{\infty} u^{\frac{\nu}{2}-1} e^{-u}\, du$$

$$= \frac{1}{\Gamma(\frac{\nu}{2})} \underbrace{\Gamma\left(\frac{\chi^2}{2}, \frac{\nu}{2}\right)}_{\text{Incomplete Gamma function}}$$

Gamma function:

$$\Gamma(a) = \int_0^{\infty} du\, e^{-u} u^{a-1} \qquad a > 0$$

Incomplete Gamma function:

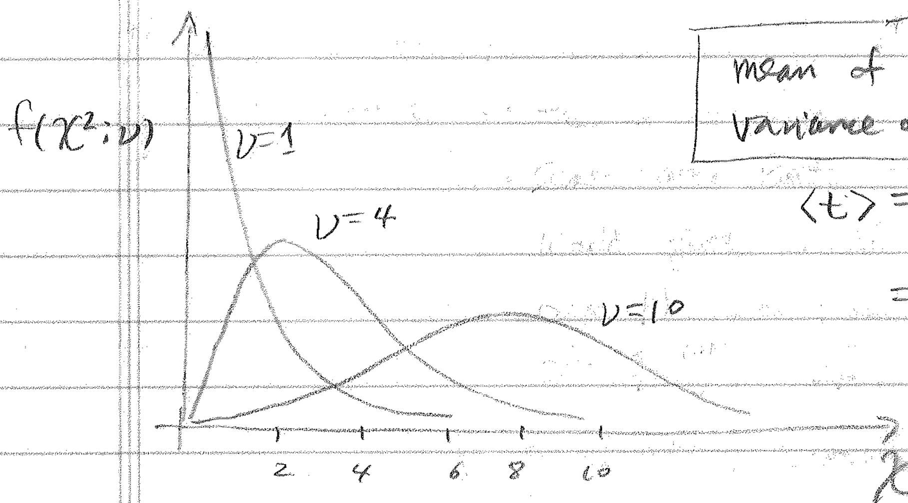
$$\Gamma(x, a) = \int_x^{\infty} e^{-u} u^{a-1}\, du \qquad x, a > 0$$

$Q(\chi^2;\nu)$: Probability that the Chi-square for a set of measurement errors obtained by randomly sampled N observations from Normal distributions would exceed the value χ^2 obtained by fitting our N data points to $\{y(x_i; a_1, \dots, a_M)\}$.

Here $\{a_1, \cdots, a_M\}$ are already determined.

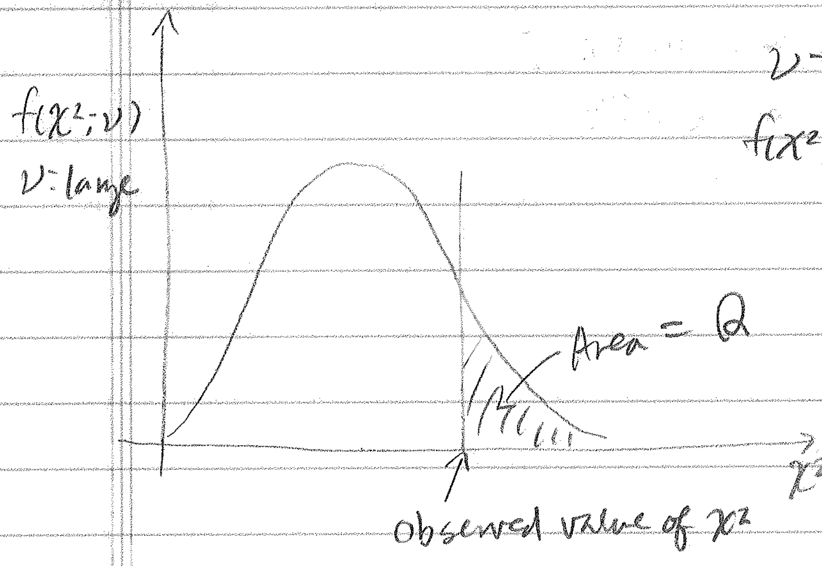
$$0 \le Q \le 1$$

$Q(\chi^2;\nu)$: quantitative measure for goodness-of-fit



mean of $f(\chi^2;\nu) = \nu$
variance of $f(\chi^2;\nu) = 2\nu$

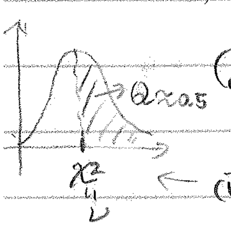
$$\langle t\rangle = \int_0^\infty t\, f(t;\nu)\, dt$$
$$= \frac{1}{2^{\nu/2}\Gamma(\nu/2)}\int_0^\infty t^{\frac{\nu}{2}} e^{-t/2}\, dt$$
$$= \nu$$



$\nu \to$ large

$f(\chi^2;\nu) \to$ Normal density function

If our model is reasonable, it is unlikely that Q is too small ($< 10^{-3}$) or too close to 1.



(i) $\boxed{Q \approx 0.5}$: Since $\nu \to$ large, $f(\chi^2;\nu) \to$ normal density ft (symmetric)

ν = mean of $f(\chi^2;\nu) \approx \chi^2$ to make the area $\frac{1}{2}$.

This is the best possible case! (Our fit is good)

But for not too large ν, $Q > 0.1$ is good enough.

$Q > 0.001$ may be acceptable.

(ii) $Q < 0.001$: fit is poor because a set of random deviations from the Normal distributions have a higher probability of giving rise to a chi-square that is equal or less than the value obtained by fitting the N data points to a model $\{y(x_i; a_1, \ldots, a_M)\}$

e.g. $Q = 0.001$: there is 99.9% probability that random deviations explain the deviation in the N data points given the model parameters.

(iii) $Q \approx 1$: $\chi^2 \approx 0$ (Too good to be true!!)

$\Rightarrow$ It is likely that standard deviations of the data $\{\sigma_i\}$ are overestimated.

- When the uncertainties associated with a set of measurements are not known in advance,

assuming $\sigma_i = \sigma$ and the model is a good fit,

$$\sigma^2 = \sum_{i=1}^{N} [y_i - y(x_i; \underbrace{a_1^{(m)}, \cdots a_M^{(m)}}_{\text{model parameters that give } \chi^2_{min}})]^2 / N$$

In this case, goodness-of-fit is not meaningful.

- $\dfrac{\partial \chi^2}{\partial a_k} = 0$, $\quad 0 = \sum_{i=1}^{N} \left(\dfrac{y_i - y(x_i)}{\sigma_i^2} \right) \left(\dfrac{\partial y(x_i; a_1, \cdots a_k, \cdots a_M)}{\partial a_k} \right)$ $\rightarrow$ M equations for M unknowns

$k = 1 \ldots M$

example 1

⟨ Fitting Data to a Straight Line ⟩

Fitting a set of N data points (x_i, y_i) to a straight-line model

$$y(x) = y(x;a,b) = a + bx$$

If the measurement errors are normally distributed, minimization of χ^2 function will give maximally likelihood estimation of the parameters a and b.

$$\chi^2(a,b) = \sum_{i=1}^{N} \left(\frac{y_i - a - bx_i}{\sigma_i} \right)^2$$

Minimization of $\chi^2(a,b)$ wrt a and b:

$$0 = \frac{\partial \chi^2}{\partial a} = -2 \sum_{i=1}^{N} \frac{y_i - a - bx_i}{\sigma_i^2}$$

$$0 = \frac{\partial \chi^2}{\partial b} = -2 \sum_{i=1}^{N} \frac{x_i(y_i - a - bx_i)}{\sigma_i^2}$$

Defining the following sums,

$$S \equiv \sum_{i=1}^{N} \frac{1}{\sigma_i^2}, \quad S_x \equiv \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}, \quad S_y \equiv \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2}, \quad S_{xx} = \sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2}, \quad S_{xy} = \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2},$$

we can rewrite the above two equations as

$$\boxed{\begin{aligned} aS + bS_x &= S_y \\ aS_x + bS_{xx} &= S_{xy} \end{aligned}}$$

What if solving this eq. is not easy by hand.

Solution of the two equations:

$$a = \frac{1}{\Delta}(S_{xx} S_y - S_x S_{xy})$$ where $\Delta = S S_{xx} - (S_x)^2$

$$b = \frac{1}{\Delta}(S S_{xy} - S_x S_y)$$

Estimate uncertainties in estimates of a and b.
(caused by measurement errors in the data)

If the data are independent,
using propagation of errors

$$\sigma_f^2 = \sum_{i=1}^{N} \sigma_{y_i}^2 \left(\frac{\partial f}{\partial y_i}\right)^2, \quad f(y_i)$$

$$\sigma_a^2 = \sum_{i=1}^{N} \sigma_i^2 \left(\frac{\partial a}{\partial y_i}\right)^2 = \sum_{i=1}^{N} \sigma_i^2 \left(\frac{S_{xx} - S_x x_i}{\sigma_i^2 \Delta}\right)^2$$

$$= \frac{1}{\Delta^2} \sum_{i=1}^{N} \frac{S_{xx}^2 - 2 S_{xx} S_x x_i + S_x^2 x_i^2}{\sigma_i^2}$$

$$= \frac{1}{\Delta^2}\left(S_{xx}^2 S - 2 S_{xx} S_x^2 + S_x^2 S_{xx}\right)$$

$$= \frac{1}{\Delta^2}\left(S_{xx}^2 S - S_x^2 S_{xx}\right) = \frac{S_{xx}}{\Delta^2}\left(S_{xx} S - S_x^2\right)$$

$$= \frac{S_{xx}}{\Delta}$$

$$\sigma_b^2 = \sum_{i=1}^{N} \sigma_i^2 \left(\frac{\partial b}{\partial y_i}\right)^2 = \sum_{i=1}^{N} \sigma_i^2 \left(\frac{S x_i - S_x}{\sigma_i^2 \Delta}\right)^2$$

$$= \frac{S}{\Delta}$$

May-13

$$\mathrm{cov}\{a,b\} \xleftarrow{\langle ab\rangle - \langle a\rangle\langle b\rangle} = -\frac{S_x}{\Delta}$$

Correlation coefficient $r_{ab} = \dfrac{\mathrm{cov}\{a,b\}}{\sigma_a \sigma_b} = \dfrac{-S_x}{\sqrt{S_{xx} S}}$

$$\begin{aligned} aS + bS_x &= S_y \\ aS_x + bS_{xx} &= S_{xy} \end{aligned} \iff \begin{bmatrix} S & S_x \\ S_x & S_{xx} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} S_y \\ S_{xy} \end{bmatrix}$$

$$[\alpha][a] = [s]$$

$$[\alpha]^{-1} = \frac{1}{\Delta}\begin{bmatrix} S_{xx} & -S_x \\ -S_x & S \end{bmatrix} = \begin{bmatrix} \sigma_a^2 & \mathrm{cov}\{a,b\} \\ \mathrm{cov}\{a,b\} & \sigma_b^2 \end{bmatrix}$$

If the parameters (a,b) are drawn from a multivariate normal distribution

$$f(a,b) = N \exp\left[-\frac{1}{2}(a \; b)\underbrace{\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}}_{[\alpha]}\begin{pmatrix} a \\ b \end{pmatrix}\right]$$

$$e^{-\frac{x^2}{2\sigma^2}}$$

why $[\alpha]$?

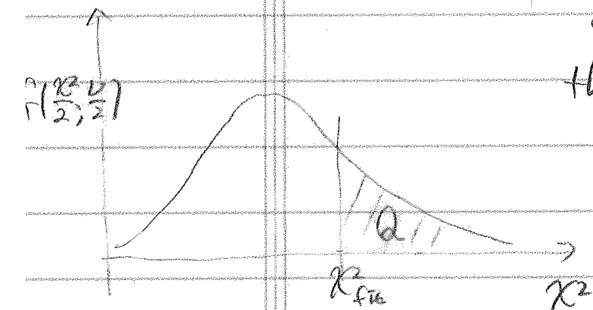
No lecture 2/16/06 (Thur) Feb 20, 06 Mon-17

Mon (2/20/06) 7:30 pm Rokeach 316

- Goodness-of-fit of the data to the model

$\nu = N - 2$

$Q\left(\frac{\nu}{2}, \frac{\chi^2}{2}\right)$: prob that random deviations from Normal distributions would provide a larger chi-square than χ^2 obtained from fitting



When you don't know σ_i in advance,

$\boxed{\text{Example 2}}$

⟨ General $\boxed{\text{Linear}}$ Least Squares ⟩

linear dependence on parameters a_k

Fitting a set of data points (x_i, y_i) to a model $y(x)$

$$y(x) = \sum_{k=1}^{M} a_k \underbrace{X_k(x)}_{\text{polynomials, sine, cosine, etc.}}$$

where $X_k(x)$: arbitrary fixed function of x, can be nonlinear functions of x

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - \sum_{k=1}^{M} a_k X_k(x_i)}{\sigma_i} \right]^2$$

Let $A_{ij} = \dfrac{X_j(x_i)}{\sigma_i}$ $\quad i = 1, \cdots, N$, $\; j = 1, \cdots, M$

$b_i = \dfrac{y_i}{\sigma_i}$, $[a] = \begin{pmatrix} a_1 \\ \vdots \\ a_M \end{pmatrix}$

Two ways to minimize χ^2 :

(I) Solve M equations for M unknowns using Gaussian Elimination or LU decomposition.

$\dfrac{\partial \chi^2}{\partial a_k} = 0, \; k = 1, \cdots, M$

(II) Find $[a]$ that minimizes $\chi^2 = \left| [A]\cdot[a] - [b] \right|^2$ using SVD.

(i) $0 = \frac{\partial \chi^2}{\partial a_k} \Rightarrow 0 = \sum_{i=1}^{N} \frac{1}{\sigma_i^2}\left[y_i - \sum_{j=1}^{M} a_j X_j(x_i) \right] X_k(x_i)$

$k = 1, \ldots, M$

$$\Rightarrow \sum_{j=1}^{M} a_j \sum_{i=1}^{N} \frac{X_j(x_i) X_k(x_i)}{\sigma_i^2} = \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2} X_k(x_i)$$

unknowns

$$\Rightarrow \sum_{j=1}^{M} \alpha_{kj} \boxed{a_j} = \beta_k$$

where $\alpha_{kj} = \sum_{i=1}^{N} \frac{X_j(x_i) X_k(x_i)}{\sigma_i^2} \iff [\alpha] = A^T \cdot A$

$$\beta_k = \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2} X_k(x_i) \iff [\beta] = A^T \cdot b$$

$$\Rightarrow [\alpha] \cdot [a] = [\beta] \quad \text{or} \quad A^T \cdot A \cdot [a] = A^T \cdot b$$

Now $[a]$ can be solved using Gaussian Elimination or LU decomposition.

Error estimate on the parameters $[a]$:

$$\sigma_{a_j}^2 = \sum_{i=1}^{N} \sigma_i^2 \left(\frac{\partial a_j}{\partial y_i} \right)^2$$

Nov-19

To find $\frac{\partial a_j}{\partial y_i}$

From $\sum_{j=1}^{M} \alpha_{kj} a_j = \beta_k$,

$$\sum_{k=1}^{M} \sum_{j=1}^{M} [\alpha]^{-1}_{\ell k} \alpha_{kj} \, a_j = \sum_{k=1}^{M} [\alpha]^{-1}_{\ell k} \beta_k$$

$$\sum_{j=1}^{M} \delta_{\ell j} a_j = a_\ell = \sum_{k=1}^{M} [\alpha]^{-1}_{\ell k} \beta_k$$

$[\alpha]^{-1}_{\ell k} \equiv C_{\ell k}$

$$a_\ell = \sum_{k=1}^{M} C_{\ell k} \left[\sum_{i=1}^{N} \frac{y_i X_k(x_i)}{\sigma_i^2} \right]$$

definition of β_k

$$\Rightarrow \frac{\partial a_j}{\partial y_i} = \sum_{k=1}^{M} \frac{C_{jk} X_k(x_i)}{\sigma_i^2}$$

$$\Rightarrow \sigma_{a_j}^2 = \sum_{i=1}^{N} \sigma_i^2 \sum_{k=1}^{M} \frac{C_{jk} X_k(x_i)}{\sigma_i^2} \sum_{\ell=1}^{M} \frac{C_{j\ell} X_\ell(x_i)}{\sigma_i^2}$$

$$= \sum_{k=1}^{M} \sum_{\ell=1}^{M} C_{jk} C_{j\ell} \underbrace{\left[\sum_{i=1}^{N} \frac{X_k(x_i) X_\ell(x_i)}{\sigma_i^2} \right]}_{\alpha_{k\ell} = [C]^{-1}_{k\ell}}$$

$$= \sum_{\ell=1}^{M} \delta_{j\ell} C_{j\ell} = C_{jj}$$

$\Rightarrow$ [Diagonal elements of $\overset{[\alpha]^{-1}}{[C]}$ = variances of fitting parameters $[a]$

Off-diagonal elements of $\underset{[\alpha]^{-1}}{[C]}$ = covariance between a_j and a_k

(II) Find $[a]$ that minimizes $\chi^2 = |[A]\cdot[a] - [b]|^2$

Let $\overset{N\times M}{A} = \overset{N\times M}{U}\ \overset{M\times M}{W}\ \overset{M\times M}{V^T}$

$U_{(i)}$: columns of U, $i=1, \dots, M$

$V_{(i)}$: " of V, $i=1, \dots, M$

$$\Rightarrow [a] = V \tilde{W}^{-1} U^T b$$

modified inverse of W

if $w_j = 0$, the correspondy element in $\tilde{W}^{-1}$ becomes zero.

$$= \sum_{i=1}^{M} \left(\frac{U_{(i)} \cdot b}{\tilde{w}_i} \right) V_{(i)}$$

by definition: $b_i = \frac{y_i}{\sigma_i}$

$$\sigma_{a_j}^2 = \sum_{i=1}^{N} \sigma_i^2 \left(\frac{\partial a_j}{\partial y_i} \right)^2$$

$$a_j = \sum_{i=1}^{M} \sum_{\ell=1}^{M} \frac{U_{\ell i} b_\ell}{w_i} V_{ji}$$

$$\frac{\partial a_j}{\partial y_i} = \sum_{k=1}^{N} \frac{\partial a_j}{\partial b_k} \cdot \frac{\partial b_k}{\partial y_i} = \sum_{k=1}^{N} \sum_{m=1}^{M} \frac{U_{km}}{w_m} V_{jm} \cdot \frac{1}{\sigma_k} \delta_{ki}$$

$$= \sum_{m=1}^{M} \frac{U_{im}}{w_m} V_{jm} \frac{1}{\sigma_i}$$

$$\sigma_{a_j}^2 = \sum_{i=1}^{N} \sigma_i^2 \left(\sum_{k=1}^{M} \frac{U_{ik}}{w_k} V_{jk} \frac{1}{\sigma_i} \right) \left(\sum_{\ell=1}^{M} \frac{U_{i\ell}}{w_\ell} V_{j\ell} \frac{1}{\sigma_i} \right)$$

$\delta_{k\ell}$

$$= \sum_{i=1}^{N} \sum_{k=1}^{M} \sum_{\ell=1}^{M} \frac{U_{ik} U_{i\ell}}{w_k w_\ell} V_{jk} V_{j\ell}$$

$$= \sum_{k=1}^{M} \left(\frac{V_{jk}}{w_k} \right)^2$$

$$\sigma_{a_j}^2 = \sum_{k=1}^{M} \left(\frac{V_{jk}}{w_k}\right)^2 \quad \& \quad cov\{a_j, a_k\} = \sum_{i=1}^{M} \left(\frac{V_{ji} V_{ki}}{w_i^2}\right)$$

can be shown from $[\alpha]^{-1}_{jl} = \sum_{i=1}^{M} \frac{V_{ji} V_{li}}{w_i^2}$

(proof) $[\alpha] = A^T A$

By definition, $[\alpha] V_{(i)} = w_i^2 V_{(i)}$ $\Leftarrow$ (from SVD)

$$[\alpha]^{-1}[\alpha] V_{(i)} = [\alpha]^{-1} w_i^2 V_{(i)}$$

$$V_{(i)} = [\alpha]^{-1} w_i^2 V_{(i)}$$

$$V_{ji} = \sum_{k=1} [\alpha]^{-1}_{jk} w_i^2 V_{ki}$$

Multiply by $\sum_{i=1}^{M} \frac{V_{li}}{w_i^2}$

$$\sum_{i=1}^{M} \frac{V_{ji} V_{li}}{w_i^2} = \sum_{k=1}^{M} \sum_{i=1}^{M} [\alpha]^{-1}_{jk} V_{ki} V_{li}$$

$$= \sum_{k=1}^{M} \delta_{kl} [\alpha]^{-1}_{jk} = [\alpha]^{-1}_{jl}$$

To avoid zero or very small pivot elements or singular matrix [A], use SVD to determine the model parameters {a}.

Drawback : Requires to store an $N \times M$ array for [A].
This storage is overwritten by matrix U.

$\boxed{\text{Example 3}}$

⟨ Nonlinear Models ⟩

Fitting a set of N data points to a model that depends nonlinearly on the set of M unknown parameters $\{a_k;\ k=1,\cdots,M\}$

With nonlinear dependence, the minimization of χ^2 must proceed iteratively because

$$\frac{\partial^2 \chi}{\partial a_k} = 0$$ does not produce linear equations in $\{a\}$.

Start with trial values for the parameters.
Improve the trial solution iteratively until χ^2 stops decreasing.

$[a] = (a_1, \cdots, a_n)$

- If we are close to the minimum of χ^2,

$$\chi^2([a]) = \gamma - [d][a] + \frac{1}{2}[a]^T[D][a]$$

$[d]$: M vector

$[D]$ = M × M matrix $\sim \nabla^2 \chi^2$

$$\vec{\nabla}\chi^2([a]) = -[d] + [D]\cdot[a]$$

at minimum $\vec{\nabla}\chi^2[a_{min}] = 0 = -[d] + [D]\cdot[a]_{min}$

If a_{cur} is close to a_{min}, $\vec{\nabla}\chi^2[a_{cur}] = -[d] + [D][a_{cur}]$

close to the minimum, if slope changes rapidly, then δa = small, otherwise δa : large.

$$[D]\cdot([a_{min}] - [a_{cur}]) = -\vec{\nabla}\chi^2[a_{cur}]$$

$$\Rightarrow [a_{min}] = [a_{cur}] + [D]^{-1}\cdot(-\vec{\nabla}\chi^2[a_{cur}]) \quad \cdots\cdots ①$$

- If we are far from the minimum of χ^2,

$$[a_{next}] = [a_{cur}] - \text{constant} \times \vec{\nabla}\chi^2[a_{cur}] \quad \cdots\cdots ②$$

where constant ($\to 0$) : small enough not to exhaust downhill direction

Since $\chi^2(a)$ is far away from the minimum only thing this does is to move a_{next} right (increase) [(left) (decrease)] if a_{cur} is on the left [(right)] hand side of the minimum.

- Calculation of the Gradient and Hessian ($\to$ second derivative)

Model to be fitted : $y = y(x; [a])$

$$\text{Minimize } \chi^2([a]) = \sum_{i=1}^{N}\left[\frac{y_i - y(x_i;[a])}{\sigma_i}\right]^2$$

$$\frac{\partial\chi^2}{\partial a_k} = -2\sum_{i=1}^{N}\frac{[y_i - y(x_i;[a])]}{\sigma_i^2}\frac{\partial y(x_i;[a])}{\partial a_k}, \quad k = 1, 2, \cdots, M$$

$$\frac{\partial^2\chi^2}{\partial a_k\,\partial a_l} = 2\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\left[\frac{\partial y(x_i;[a])}{\partial a_k}\frac{\partial y(x_i;[a])}{\partial a_l}\right.$$

$$- \underbrace{[y_i - y(x_i;[a])]}_{\uparrow} \frac{\partial^2 y_i(x_i;[a])}{\partial a_l \partial a_k} \Bigg]$$

for a successful model, this will be the random measurement error of each point. Σ I. cancel out.

$$\approx 2 \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \frac{\partial y(x_i;[a])}{\partial a_k} \cdot \frac{\partial y(x_i;[a])}{\partial a_l}$$

Define

$$\beta_k \equiv -\frac{1}{2}\frac{\partial \chi^2}{\partial a_k}, \quad \alpha_{kl} \equiv \frac{1}{2}\frac{\partial^2 \chi^2}{\partial a_k \partial a_l}$$

Eq ① : $[\delta a] \underbrace{\boxed{[D]\frac{1}{2}}}_{[\alpha]} = \underbrace{\frac{1}{2}(-\vec{\nabla}\chi^2[a_{cur}])}_{[\beta]}$

$$\Rightarrow \sum_{l=1}^{M} \alpha_{kl}\, \delta a_l = \beta_k \quad \cdots \quad ③$$

<Levenberg - Marquardt Method>

Essence : If $\chi^2[a]$ is far away from the minimum, then use Eq. ② to update $[a]$. Then once $\chi^2[a]$ becomes close to the minimum, use Eq. ① to update $[a]$.

Eq. ② : To get a sense of the constant, use dimensional analysis.

$D^{-1} = \alpha^{-1}$

const $\sim \dfrac{1}{\lambda \alpha_{ll}}$ $\quad \lambda$ = fudge factor

$$\delta a_\ell = \left(\frac{1}{\lambda \alpha_{\ell\ell}}\right)\beta_\ell \quad \text{or} \quad \lambda \alpha_{\ell\ell}\, \delta a_\ell = \beta_\ell \quad \cdots \; ④$$ Mar-25

To do this, (use Eq. ② and ①) modify Eq ③ to

$$\sum_{\ell=1}^{M} \alpha'_{k\ell}\, \delta a_\ell = \beta_k \; ; \quad \alpha'_{jj} = \alpha_{jj}(1+\lambda) \quad \cdots\cdots ⑤$$

$$\alpha'_{jk} \equiv \alpha_{jk} \; (j \neq k)$$

$$\begin{cases} \lambda : \text{large} \Rightarrow \text{Eq. ② (Eq. ④)} \\ \lambda : \text{zero} \Rightarrow \text{Eq. ③ (Eq. ①)} \end{cases}$$

- Algorithm for Levenberg-Marquardt Method

(i) Compute $\chi^2([a])$.

(ii) Pick a modest value for λ, say $\lambda = 0.001$.

(iii) Solve Eq. ⑤ for $[\delta a]$ and evaluate $\chi^2([a+\delta a])$.

(iv) If $\chi^2([a+\delta a]) \geq \chi^2([a])$, increase λ by a factor of 10 and go back to (iii).

We are far away from the minimum

If $\chi^2([a+\delta a]) < \chi^2([a])$, decrease λ by a factor of 10, update the trial solution $a+\delta a$ to a and go back to (iii).

(v) Stop iteration if χ^2 decreases by a negligible amount first time or second time.

(vi) When χ^2 reaches the minimum, set $\lambda = 0$ and compute the matrix $[C] \equiv [\alpha]^{-1}$ to get the variance and covariances of the model parameters.