Systematics, Correlations and Combination

Louis Lyons

Imperial College and Oxford

Goettingen October 2010

With thanks to Arnulf and Thomas

Systematics

J. Heinrich and L. Lyons, Ann Rev of Nucl + Particle Science 57 (2007) 145

Lots of information on Systematics (and other topics) on CDF Statistics Committee web-site

Random + Systematic Errors

Random/Statistical: Limited accuracy, Poisson counts

Spread of answers on repetition (Method of estimating)

Systematics: May cause shift, but not spread

e.g. Pendulum $g = 4\pi^2 L/\tau^2$, $\tau = T/n$

Statistical errors: T, L

Systematics: T, L

Calibrate: Systematic → Statistical

More systematics:

Formula for **undamped, small amplitude, rigid, simple** pendulum

Might want to correct to g at sea level:

Different correction formulae

Ratio of g at different locations: Possible systematics might cancel.

Correlations relevant

Presenting result

Quote result as $g \pm \sigma_{stat} \pm \sigma_{syst}$

Or combine errors in quadrature → $g \pm \sigma$

Other extreme: Show all systematic contributions separately

Useful for assessing correlations with other measurements

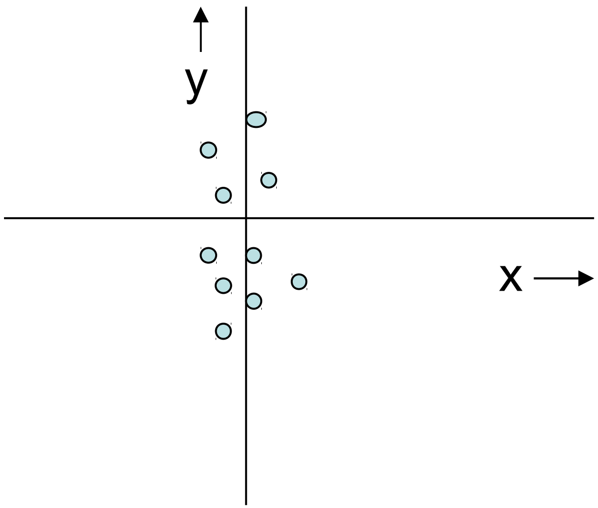
Needed for using:

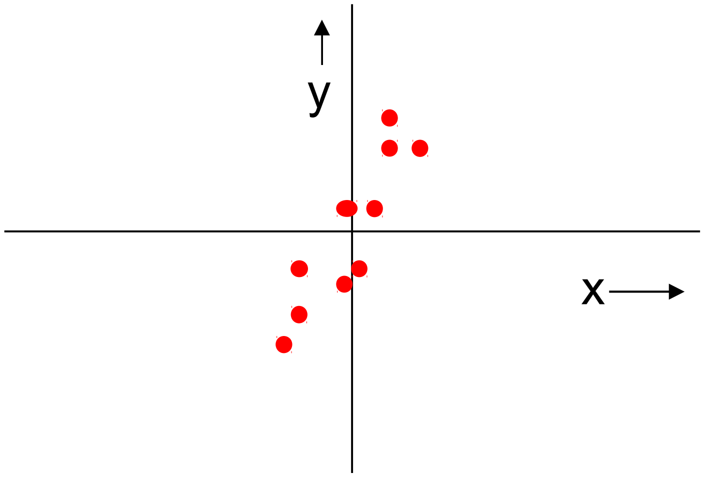
improved outside information,

combining results

using measurements to calculate something else.

Correlations and Error Matrix





Cov(x,y) ~ 0

Cov(x,y) > 0

N.B. Correlations of errors, not of variables

e.g. Period and length of pendulum.

Gaussian or Normal

$$y = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

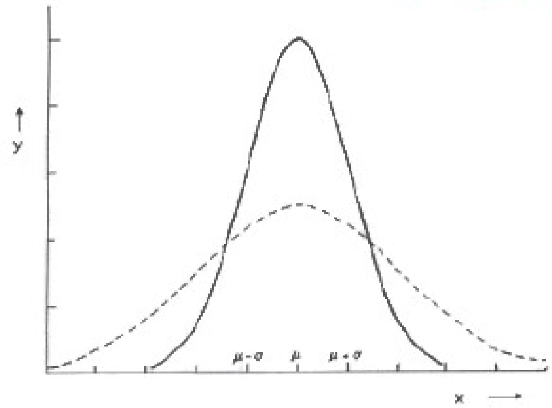


Fig. 1.5. The solid curve is the Gaussian distribution of eqn (1.14). The distribution peaks at the mean μ, and its width is characterised by the parameter σ. The dashed curve is another Gaussian distribution with the same values of μ, but with σ twice as large as the solid curve. Because the normalisation condition (1.15) ensures that the area under the curves is the same, the height of the dashed curve is only half that of the solid curve at their maxima. The scale on the x-axis refers to the solid curve.

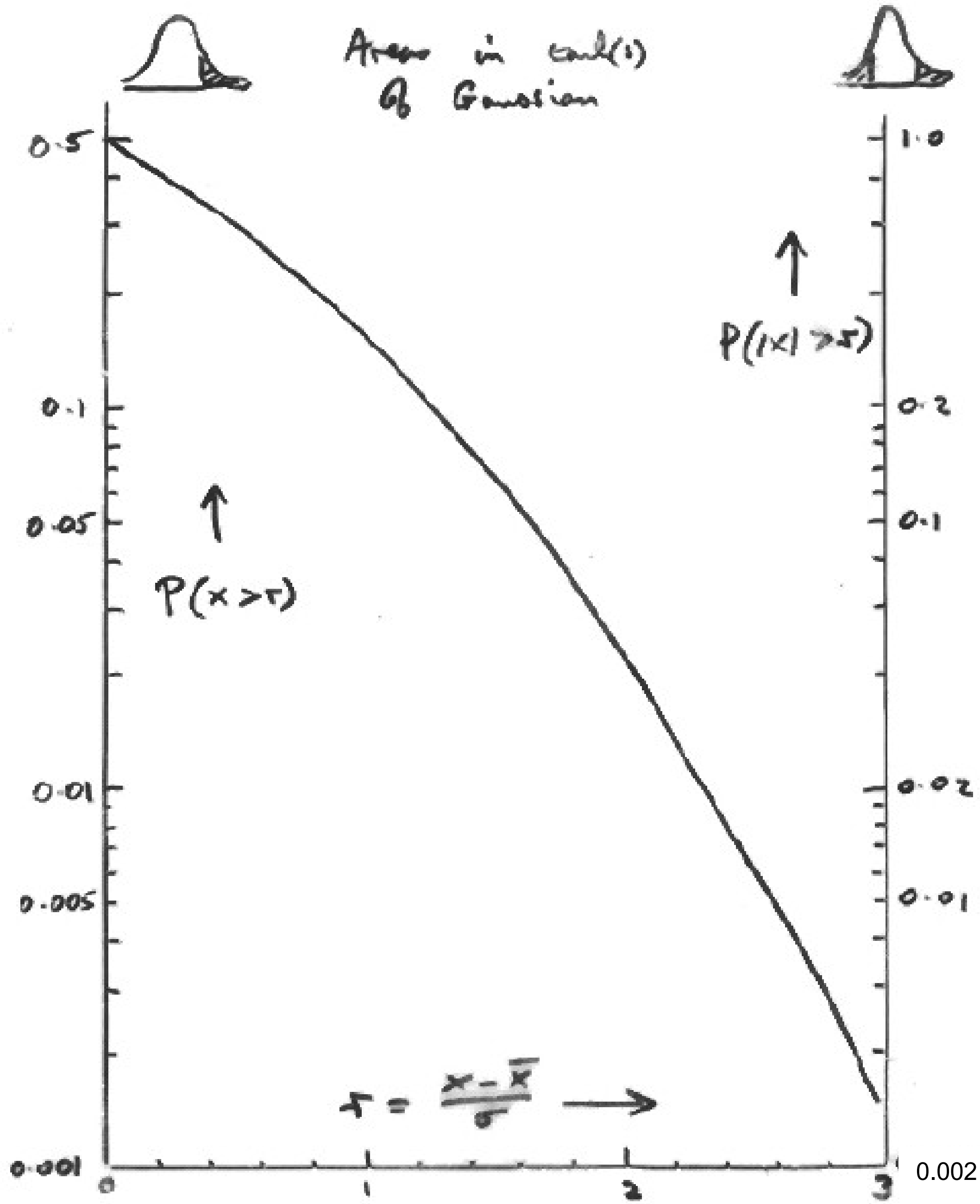
Significance of σ

i) RMS of Gaussian $= \sigma$
(Hence factor of 2 in defn of Gaussian)

ii) At $x = \mu \pm \sigma$, $y = y_{max}/\sqrt{e}$
(i.e. $\sigma \sim$ half-width at "half" height)

iii) Fractional area within $\mu \pm \sigma$ is 68%

iv) Height at max $= 1/\sqrt{2\pi}\,\sigma$



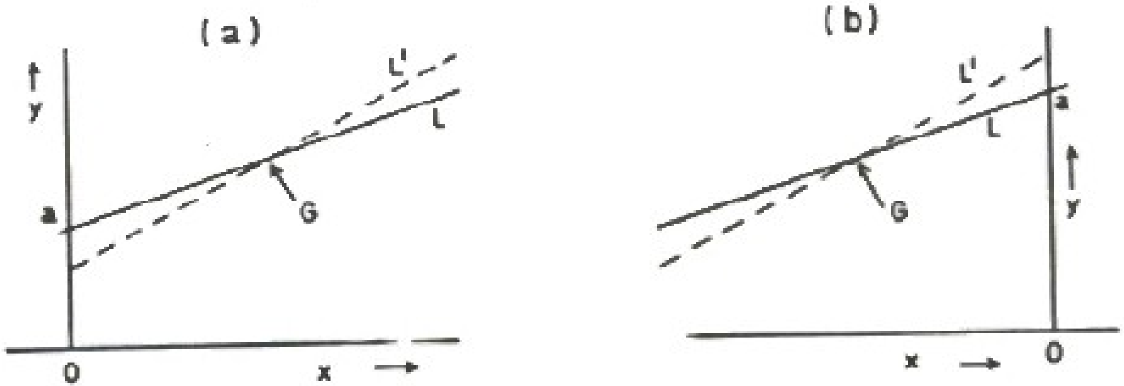
Area in tail(s)
of Gaussian
0.5
0.1
0.05
0.01
0.005
0.001
1.0
0.2
0.1
0.02
0.01
0.002
P(x > r)
P(|x| > r)
$r = \frac{x - \bar{x}}{\sigma}$
0
1
2
3

Learning to love the Error Matrix

- Introduction via 2-D Gaussian
- Understanding covariance
- Using the error matrix

  Combining correlated measurements

COVARIANCE $(a, b) \propto -\langle x \rangle$



$\langle x \rangle$ pos

Fig. 2.4

$\langle x \rangle$ neg

Correlations

Basic issue:

For 1 parameter, quote value and error

For 2 (or more) parameters,

(e.g. gradient and intercept of straight line fit)

quote values + errors **+ correlations**

Just as the concept of variance for single variable is more general than Gaussian distribution, so correlation in more variables does not require multi-dim Gaussian

But simpler to introduce concept this way

Start with 2 variables with uncorrelated errors, and then introduce correlations in a simple way

Gaussian in 2-variables

$$P(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_x} e^{-\frac{1}{2}\frac{x^2}{\sigma_x^2}}$$

$$P(y) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_y} e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}}$$

x + y uncorrelated ⇒

$$P(x,y) = \frac{1}{2\pi} \frac{1}{\sigma_x \sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)}$$

Down on P(0,0) by $e^{-\frac{1}{2}}$ when

$$\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} = 1$$

Rewrite as

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_x^2} & 0 \\ 0 & \frac{1}{\sigma_y^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1$$

Invert ⇒ ERROR MATRIX

$$\begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}$$

Element E_{ij} - $\langle (x_i - \overline{x}_i)(x_j - \overline{x}_j) \rangle$

Diagonal E_{ij} = variances

Off-diagonal E_{ij} = covariances

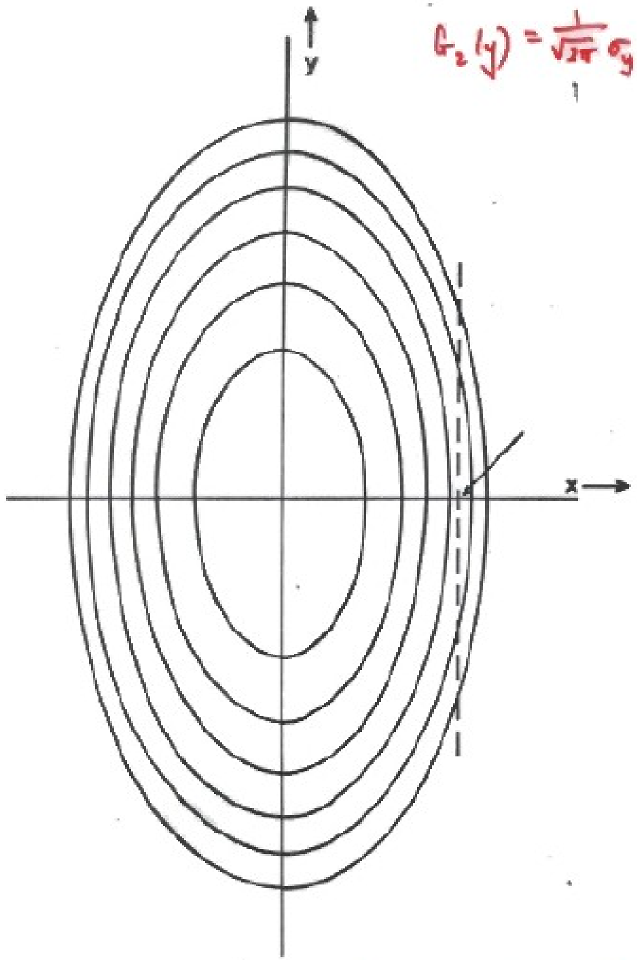
TOWARDS THE ERROR MATRIX

$x + y$ indep Gaussians

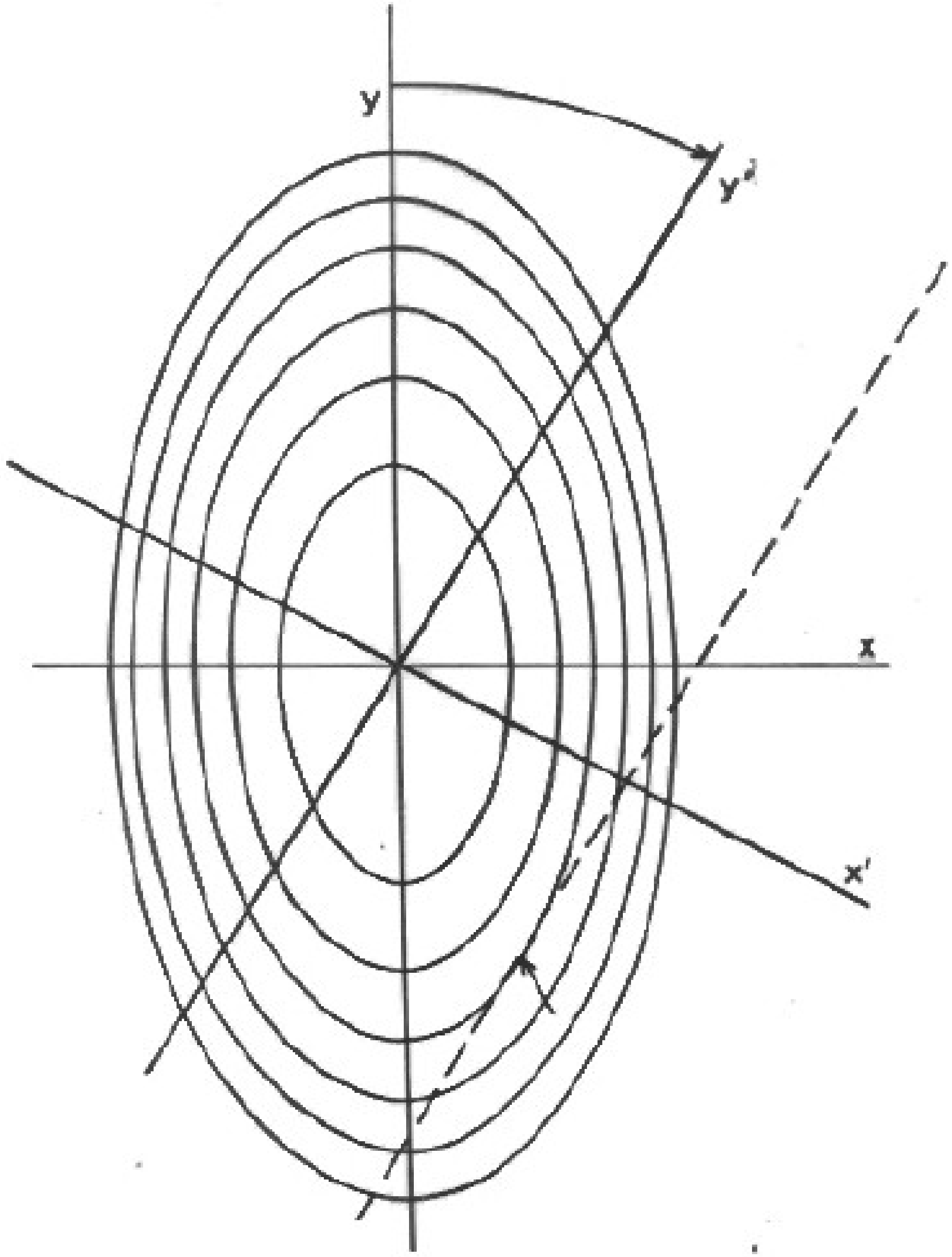
$$P(x, y) = G_1(x)\, G_2(y)$$

$$G_1(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} \exp\left[-\frac{1}{2}\frac{x^2}{\sigma_x^2}\right]$$

$$G_2(y) = \frac{1}{\sqrt{2\pi}\,\sigma_y} \exp\left[-\frac{1}{2}\frac{y^2}{\sigma_y^2}\right]$$



$$P(x, y) = \frac{1}{2\pi}\frac{1}{\sigma_x \sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right]$$


y
y'
x
x'

Specific example

$\sigma_x = \frac{\sqrt{2}}{4} = .354$ $\sigma_y = \frac{\sqrt{2}}{2} = .707$

Then factor of $e^{-\frac{1}{2}}$ when

$$8x^2 + 2y^2 = 1$$

Now introduce CORRELATIONS by 30° rotn

$$\frac{1}{2}\left[13x'^2 + 6\sqrt{3}\,x'y' + 7y'^2\right] = 1$$

$$\begin{pmatrix} \frac{13}{2} & \frac{3\sqrt{3}}{2} \\ \frac{3\sqrt{3}}{2} & \frac{7}{2} \end{pmatrix} = \text{Inverse Error Matrix}$$

$$\frac{1}{32} \times \begin{pmatrix} 7 & -3\sqrt{3} \\ -3\sqrt{3} & 13 \end{pmatrix} = \text{Error Matrix}$$

$8x^2 + 2y^2 = 1$

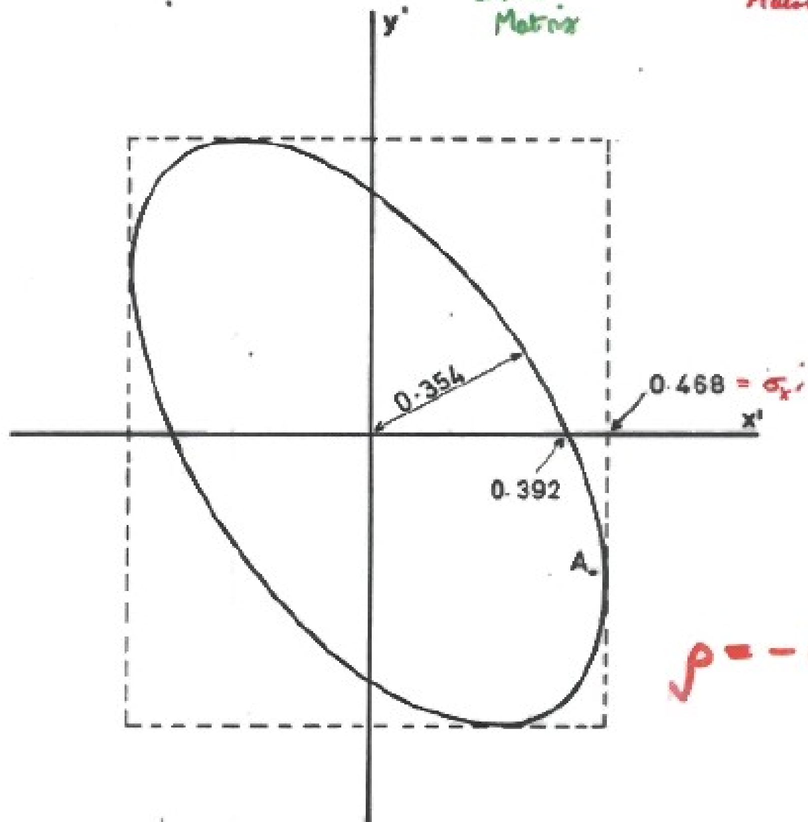
$\frac{1}{2}[13x'^2 + 6\sqrt{3}x'y' + 7y'^2] = 1$

$\begin{pmatrix} \frac{13}{2} & \frac{3\sqrt{3}}{2} \\ \frac{3\sqrt{3}}{2} & \frac{7}{2} \end{pmatrix}$

Inverse Error Matrix

$\frac{1}{32} \times \begin{pmatrix} 7 & -3\sqrt{3} \\ -3\sqrt{3} & 13 \end{pmatrix}$

Error Matrix



$\rho = -0.54$

$(0.468)^2 = \frac{7}{32} = \sigma_{x'}^2$

$(0.392)^2 = 1/6.5$

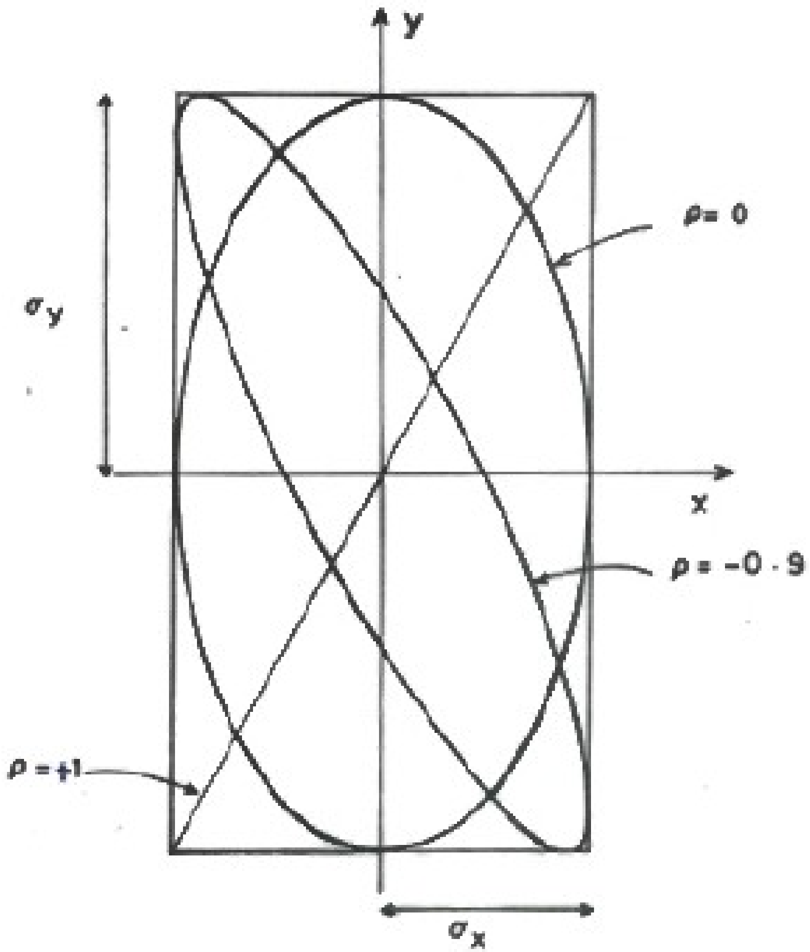
$\frac{1}{8} = (0.354)^2$ = Eigenvalue of error matrix $= \sigma_x^2$

σ_x, σ_y } constant

ρ varying

Covariance → $\begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$

Error Matrix



USING THE ERROR MATRIX

(i) Function of variables

$$y = y(x_a, x_b)$$

Given x_a, x_b error matrix, what is σ_y?

Differentiate, square, average

$$\overline{\delta y^2} = \left(\frac{\partial y}{\partial x_a}\right)^2 \overline{\delta x_a^2} + \left(\frac{\partial y}{\partial x_b}\right)^2 \overline{\delta x_b^2} + 2\frac{\partial y}{\partial x_a}\frac{\partial y}{\partial x_b}\overline{\delta x_a \delta x_b}$$

Zero if x_a, x_b uncorrelated

OR

$$\overline{\delta y^2} = \begin{pmatrix}\frac{\partial y}{\partial x_a} & \frac{\partial y}{\partial x_b}\end{pmatrix} \begin{pmatrix}\overline{\delta x_a^2} & \overline{\delta x_a \delta x_b} \\ \overline{\delta x_a \delta x_b} & \overline{\delta x_b^2}\end{pmatrix} \begin{pmatrix}\frac{\partial y}{\partial x_a} \\ \frac{\partial y}{\partial x_b}\end{pmatrix}$$

$\tilde{D}$ — Error matrix — Derivative vector D

$$\sigma_y^2 = \tilde{D} E D$$

(ii) Change of variables $\quad x_a = x_a(p_i, p_j)$
$x_b = x_b(p_i, p_j)$

e.g. Cartesian $\Rightarrow$ polars

or Points in $x, y \Rightarrow m, c$ of straight line fit

Given (p_i, p_j) error matrix $\Rightarrow (x_i, x_j)$ error matrix

Differentiate, $\delta x_a \delta x_b$, average

$$\delta x_a = \frac{\partial x_a}{\partial p_i}\delta p_i + \frac{\partial x_a}{\partial p_j}\delta p_j \qquad (+\text{ sim for } x_b)$$

Then

$$\overline{\delta x_a^2} = \left(\frac{\partial x_a}{\partial p_i}\right)^2 \overline{\delta p_i^2} + \left(\frac{\partial x_a}{\partial p_j}\right)^2 \overline{\delta p_j^2} + 2\frac{\partial x_a}{\partial p_i}\frac{\partial x_a}{\partial p_j}\overline{\delta p_i \delta p_j}$$

$$\overline{\delta x_a \delta x_b} = \frac{\partial x_a}{\partial p_i}\frac{\partial x_b}{\partial p_i}\overline{\delta p_i^2} + \frac{\partial x_a}{\partial p_j}\frac{\partial x_b}{\partial p_j}\overline{\delta p_j^2} + \left(\frac{\partial x_a}{\partial p_i}\frac{\partial x_b}{\partial p_j} + \frac{\partial x_a}{\partial p_j}\frac{\partial x_b}{\partial p_i}\right) \times \overline{\delta p_i \delta p_j}$$

$+ \overline{\delta x_b^2}$ like $\overline{\delta x_a^2}$

N.B. Change of variables does not have to be $N \to N$

e.g. straight line fit involves $N \to 2$

Then i) & ii) are both examples of $N \to M$ $(M \leq N)$

where $M = 1$ in i) $\qquad M = N$ in ii) i.e.

$$\begin{pmatrix} \overline{\delta x_a^2} & \overline{\delta x_a \delta x_b} \\ \overline{\delta x_a \delta x_b} & \overline{\delta x_b^2} \end{pmatrix} = \begin{pmatrix} \frac{\partial x_a}{\partial p_i} & \frac{\partial x_a}{\partial p_j} \\ \frac{\partial x_b}{\partial p_i} & \frac{\partial x_b}{\partial p_j} \end{pmatrix} \begin{pmatrix} \overline{\delta p_i^2} & \overline{\delta p_i \delta p_j} \\ \overline{\delta p_i \delta p_j} & \overline{\delta p_j^2} \end{pmatrix} \begin{pmatrix} \frac{\partial x_a}{\partial p_i} & \frac{\partial x_b}{\partial p_i} \\ \frac{\partial x_a}{\partial p_j} & \frac{\partial x_b}{\partial p_j} \end{pmatrix}$$

↑ New error matrix ↑ $\tilde{T}$ ↑ Old error matrix ↑ Transform matrix T

$$E_x = \tilde{T} E_p T$$

BEWARE!

e.g

$$\sigma_M^2 = \tilde{D}\,\tilde{T}\,E\,T\,D$$

Tracks' error matrix (centre of tracks)

Transformation matrix from centre of tracks to vertex

Deriv vector for mass in terms of track params at vertex

Combination

Better to combine data than to combine results

COMBINING EXPERIMENTS

$x_i \pm \sigma_i$ (uncorrelated)

$$\hat{x} = \frac{\sum x_i/\sigma_i^2}{\sum 1/\sigma_i^2}$$

$$1/\sigma^2 = \sum 1/\sigma_i^2$$

From $S = \sum (x_i - \hat{x})^2/\sigma_i^2$

Minimise S (→ $\hat{x}$)

σ from $S_{min} + 1$ (→ $1/\sigma^2$)

OR Propagate errors from $\hat{x} = \ldots$

Define $w_i = 1/\sigma_i^2$ = weight ~ information content

$$\hat{x} = \sum w_i x_i / \sum w_i$$

$$W = \sum w_i$$

Example: Equal σ_i $\Rightarrow$ $\hat{x} = \bar{x}$

$$\sigma = \sigma_i/\sqrt{n}$$

BEWARE

100 ± 10 and 1 ± 1 $\Rightarrow$ 2 ± 1 ?

or 50.5 ± 5 ?

Likelihood approach

Want to combine n_1 and n_2

Likelihood method avoids problem

$pdf(n) = e^{-\mu} \mu^n / n!$

$L(\mu) = (e^{-\mu} \mu^{n1} / n_1!) (e^{-\mu} \mu^{n2} / n_2!)$

$= e^{-2\mu} \mu^{(n1+n2)} / n_1! \, n_2!$

Maximises for $\mu = (n_1+n_2)/2$

$\sigma_\mu^2 = (n_1+n_2)/4$

e.g. 1 and 100 → 50.5 ± 5

Error estimate

N.B. Both χ^2 and likelihood approaches give uncertainty on combined result which depends on individual errors, but not on their consistency

e.g. 0 ± 3 and $2\pm3 \rightarrow 1\pm2$

0 ± 3 and $20\pm3 \rightarrow 10\pm2$

PDG have procedure to allow for spread

Cf: Errors on straight line fit

To consider……

Is it possible to combine

1 ± 10 and 2 ± 9

to get a best combined value of

6 ± 1 ?

Answer later.

Difference between averaging and adding

Isolated island with conservative inhabitants

How many married people ?

Number of married men = 100 ± 5 K

Number of married women = 80 ± 30 K

Total = 180 ± 30 K

Wtd average = 99 ± 5 K

Total = 198 ± 10 K

CONTRAST

GENERAL POINT: Adding (uncontroversial) theoretical input can improve precision of answer

Compare “kinematic fitting”

Combining correlated measurements

If $a_i \pm \sigma_i$ are independent:

Minimise $S = \sum \left(\frac{a_i - \hat{a}}{\sigma_i}\right)^2$

$$\Rightarrow \quad \hat{a} = \frac{\sum a_i w_i}{\sum w_i} \qquad w_i = 1/\sigma_i^2$$

Now $a_i \pm \sigma_i$ are correlated with error matrix $\underline{\underline{E}}$

$$\underline{\underline{E}} = \begin{pmatrix} \sigma_1^2 & cov(1,2) & cov(1,3) & \cdots \\ cov(1,2) & \sigma_2^2 & cov(2,3) & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

$$S = \sum_{i,j} (a_i - \hat{a}) \underline{\underline{E}}^{-1}_{ij} (a_j - \hat{a})$$

↑ INVERSE ERROR MATRIX

N.B. $\hat{a}$ CAN LIE OUTSIDE a_i

$\sigma_a \to 0$ AS $\rho \to \pm 1$

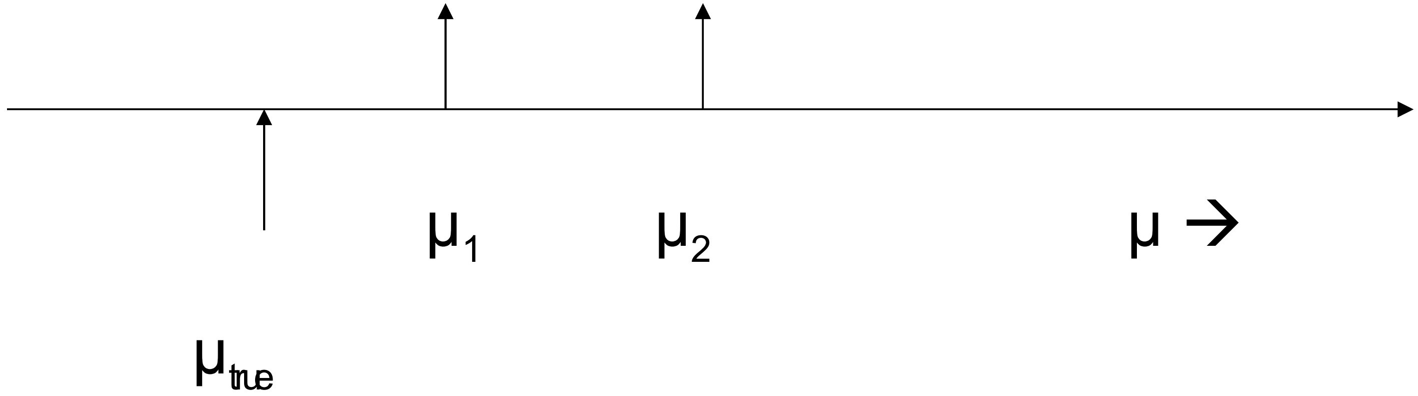
$$\underline{\underline{E}}^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 & 0 & \cdots \\ 0 & 1/\sigma_2^2 & 0 & \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

FOR UNCORRELATED

Combined value outside range of individual measurements!

$\sigma_1 < \sigma_2$ Cov > 0

Combined value outside range of individual measurements!



BLUE Lyons, Gibaut + Clifford

NIM A270 (1988) 110

Equivalent method of combining correlated measurements $a_i \pm \sigma_i$ with error matrix E_{ij}

$(E_{ii} = \sigma_i^2)$

Look for a_{best} = Best Linear Unbiassed Estimate

$a_{best} = \Sigma\, \alpha_i\, a_i$ with $\Sigma\alpha_i = 1$

Then minimise $\sigma^2 = \Sigma\alpha_i\, E_{ij}\, \alpha_j$ with respect to α's,

to get a_{best} , σ^2 and $\chi^2 = \Sigma\Sigma(a_i - a_{best})\, E_{ij}\, (a_j - a_{best})$

Because α's are known, can calculate statistical and systematic errors for a_{best}

Examples of BLUE

1) Measurement 2 is identical to meas 1

   No improvement

2) Data for meas 2 is subset of data for 1

   Meas 2 is ignored

3) General case $\sigma_1 < \sigma_2$, corrln coeff r

   $r = 0$, uncorrelated case

   $r = \sigma_2/\sigma_1$, second meas ignored

   $r > \sigma_2/\sigma_1$, second meas has negative weight

   ‘Extrapolation’

   Error→0 as r→1

MORE COMBINING:

SEVERAL PAIRS OF CORRELATED MEAS.

(x_i, y_i) with $\underline{\underline{E}}_i = \begin{pmatrix} \sigma_x^2 & cov \\ cov & \sigma_y^2 \end{pmatrix}_i$

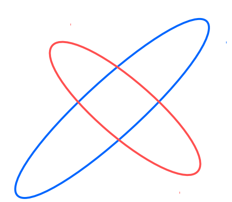
$$S = \sum_i \Big\{ (x_i - \hat{x})^2 E^{-1}_{11,i} + (y_i - \hat{y})^2 E^{-1}_{22,i} + 2(x_i - \hat{x})(y_i - \hat{y}) E^{-1}_{12,i} \Big\}$$

ice result: —

Inverse error matrix on result $\hat{x}, \hat{y}$

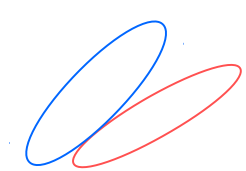
$$= \sum_i \underline{\underline{E}}_i^{-1}$$

cf $\frac{1}{\sigma^2} = \sum \frac{1}{\sigma_i^2}$ for single uncorrelated meas.



Small error

Example: Chi-sq Lecture



x_{best} outside $x_1 \rightarrow x_2$

y_{best} outside $y_1 \rightarrow y_2$

COVARIANCE (a, b) ∝ −⟨x⟩

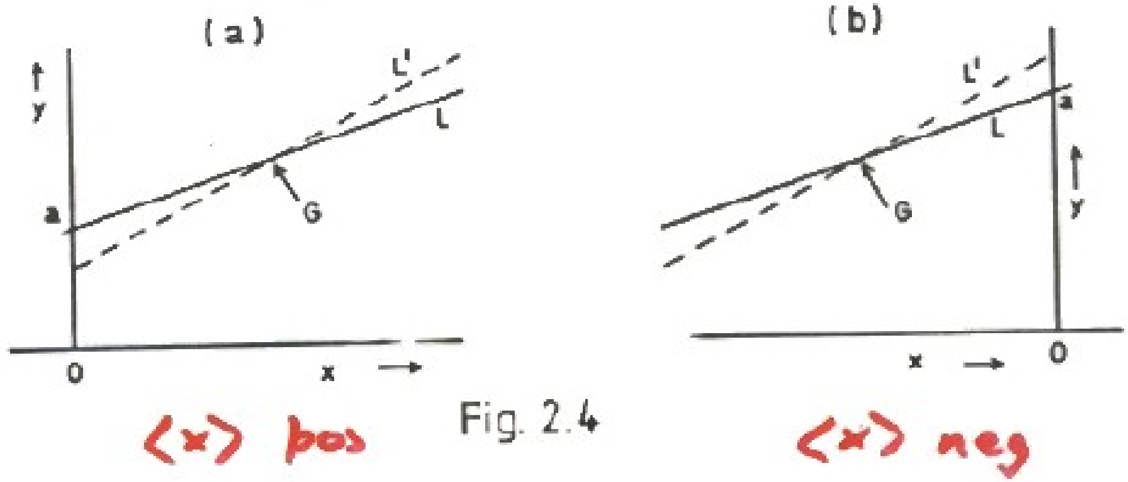
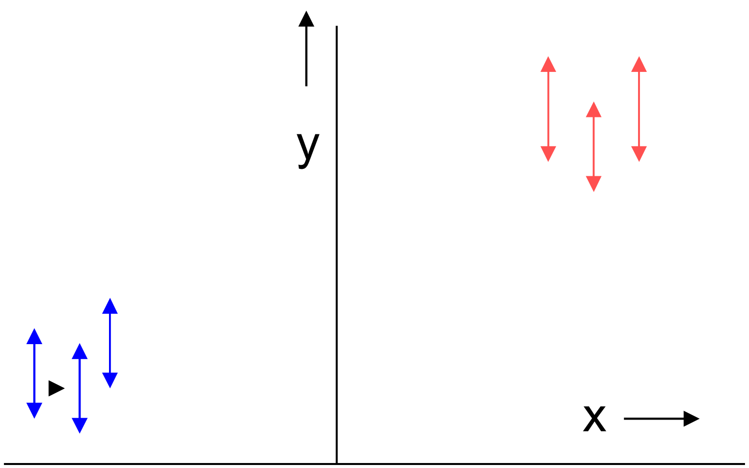
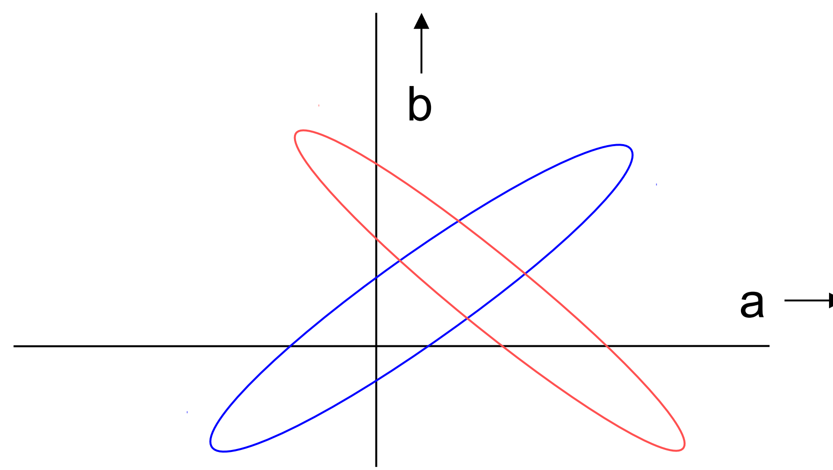
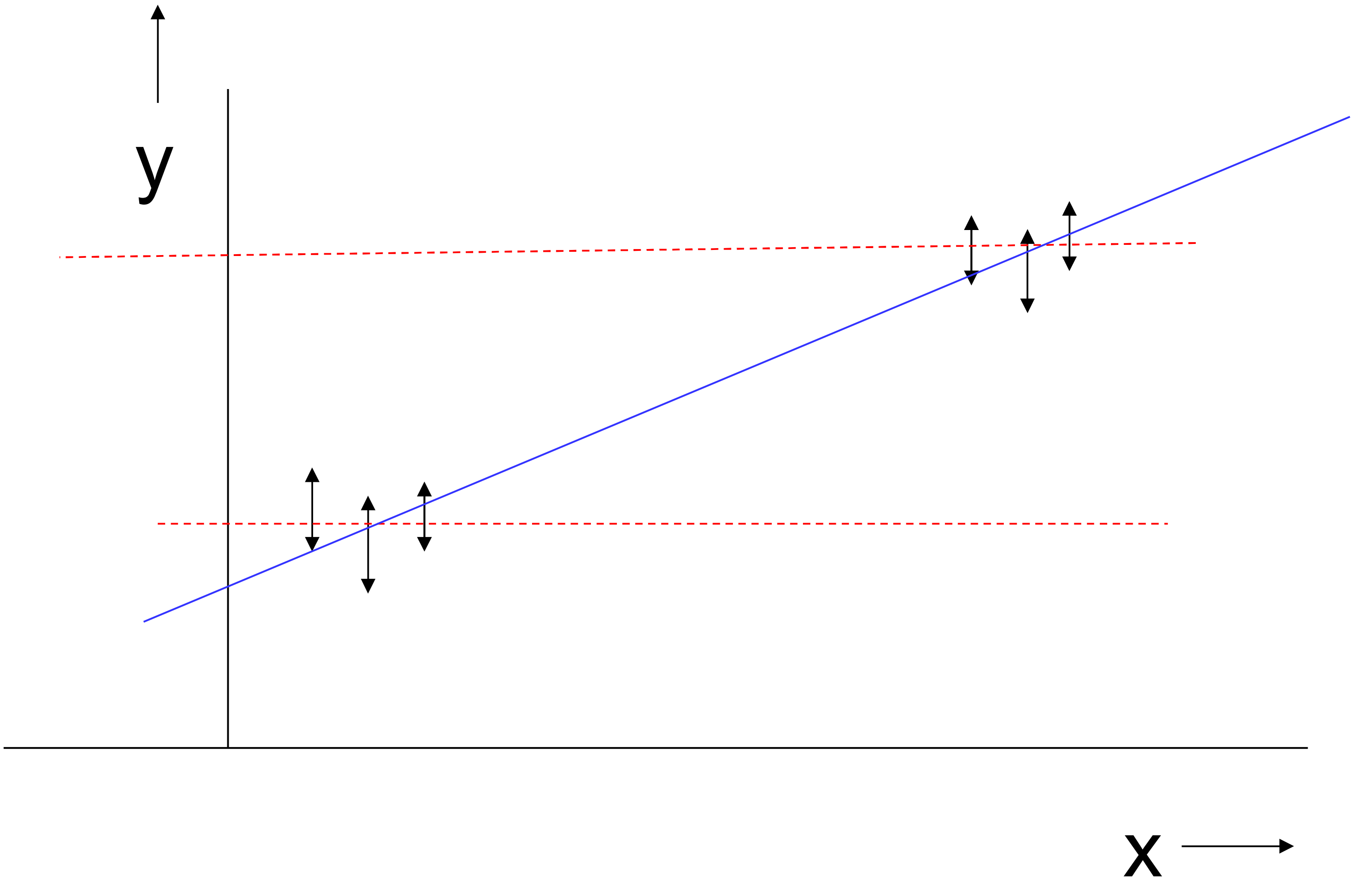


Fig. 2.4

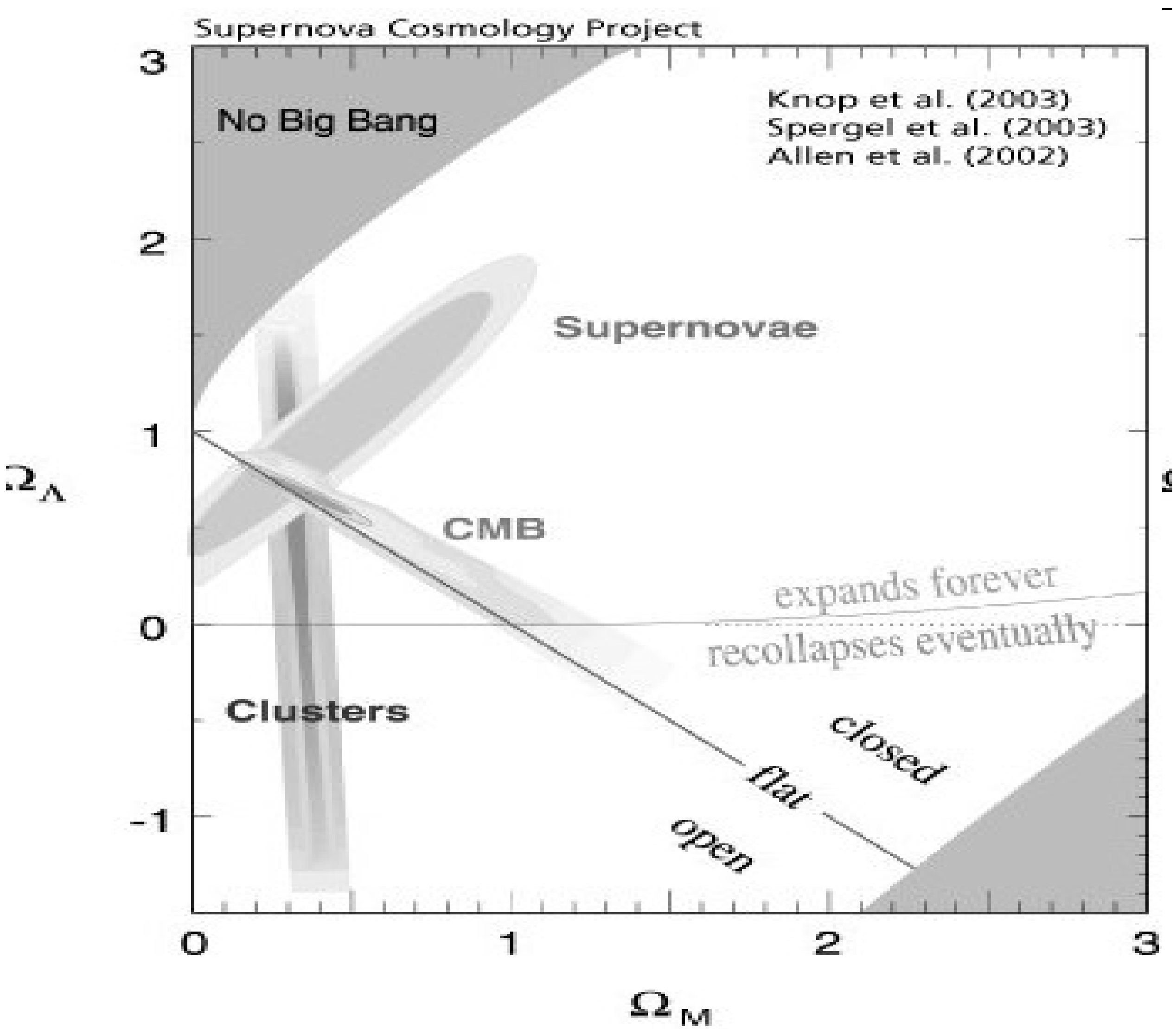




Best fit values both outside range



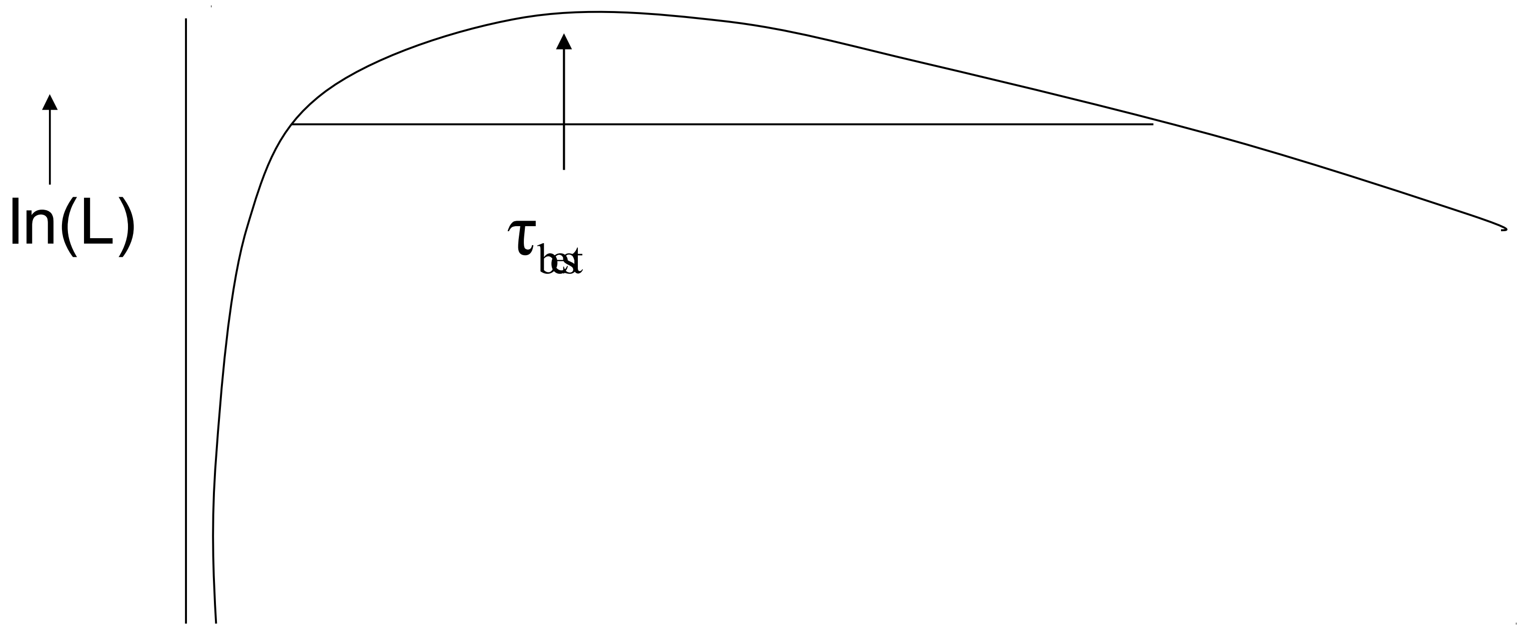
Combining error matrices in Cosmology



Non-Gaussian errors

Apart from Poisson, all above was for Gaussian errors.

Sometimes errors asymmetric e.g. lifetimes



Asymmetric errors

Barlow at:

PHYSTAT2003, page 250

PHYSTAT05, page 56

arXiv:physics/0306138v1 [physics.data-an]

18 June 2003

Error combinations, weighted sums, χ^2

BASIC PROBLEM: Meaning of asymmetric errors

Combining non-G measurements

For statistical errors, Likelihoods are great

With systematics:

INCORRECT: Combine likelihoods for statistical errors, and then find effect of systematics.

Bad for uncertainties like

$$(\sigma_{stat}, \sigma_{syst}) = (10,1) \text{ and } (1,10)$$

Better to use profile L or Bayes

Bayes

Use Bayes' theorem:

$p(\varphi,\nu|\text{data}) \sim p(\text{data}|\varphi,\nu)\ \pi(\varphi,\nu)$

Bayes posterior — Likelihood — Bayes prior

$\pi(\varphi,\nu) \sim \pi(\varphi)\ \pi(\nu)$

$\pi(\nu)$ from subsidiary measurement

$\pi(\varphi)$ from ……..

Prior knowledge better than prior ignorance.

Constant prior?

Finally integrate $p(\varphi,\nu|\text{data})$ over ν to get $p(\varphi|\text{data})$

'Marginalisation'. Contrast 'Profiling'

More in later lectures

Combining p-values

As usual, better to combine data than to combine results

Combining different p-values

Several results quote p-values for same effect: p_1, p_2, p_3…..

e.g. 0.9, 0.001, 0.3 ……..

What is combined significance? Not just $p_1 * p_2 * p_3$…..

If 10 expts each have p ~ 0.5, product ~ 0.001 and is clearly **NOT** correct combined p

$$S = z \; * \sum_{j=0}^{n-1} (-\ln z)^j / j! \; , \qquad z = p_1 p_2 p_3 \ldots\ldots.$$

(e.g. For 2 measurements, $S = z * (1 - \ln z) \geq z$)

Slight problem: Formula is not associative

Combining {{p_1 and p_2}, and then p_3} gives different answer

from {{p_3 and p_2}, and then p_1} , or all together

Due to different options for "more extreme than x_1, x_2, x_3".

Combining different p-values

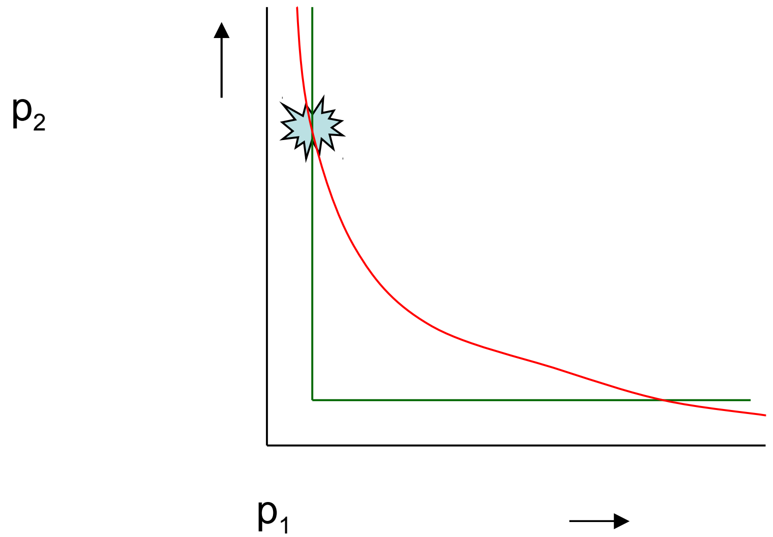
Conventional:

Are set of p-values consistent with H0?

SLEUTH:

How significant is smallest p?

$$1-S = (1-p_{smallest})^n$$



	$p_1 = 0.01$		$p_1 = 10^{-4}$	
	$p_2 = 0.01$	$p_2 = 1$	$p_2 = 10^{-4}$	$p_2 = 1$
Combined S				
Conventional	$1.0\ 10^{-3}$	$5.6\ 10^{-2}$	$1.9\ 10^{-7}$	$1.0\ 10^{-3}$
SLEUTH	$2.0\ 10^{-2}$	$2.0\ 10^{-2}$	$2.0\ 10^{-4}$	$2.0\ 10^{-4}$