χ^2 and Goodness of Fit

Louis Lyons
Imperial College & Oxford

TRISEP

June 2021

Least squares best fit

Resume of straight line

Correlated uncertainties

Uncertainties in x and in y

Goodness of fit with χ^2

Errors of first and second kind

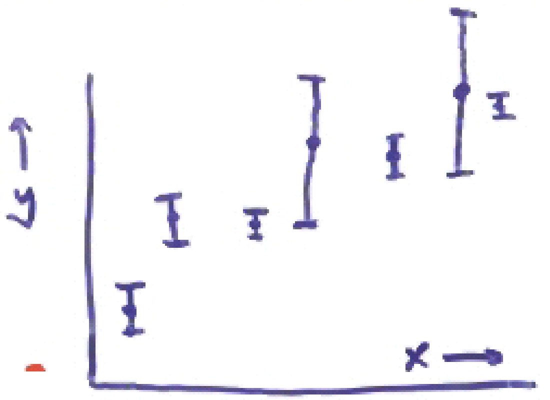
Kinematic fitting

Toy example

THE paradox

Least Squares Straight Line Fitting

Data = $\{x_i, y_i \pm \delta y_i\}$



1) Does it fit straight line?

(Goodness of Fit)

2) What are gradient and intercept?

(Parameter Determination)

Do 2) first

N.B.1 Can be used for non "a+bx"

e.g. $a + b/x + c/x^2$

N.B.2 Least squares is not the only method

$$S = \sum_i \left(\frac{y_i^{th} - y_i^{obs}}{\sigma_i} \right)^2$$

σ_i is supposed to be ‘uncertainty on data if it agreed with theory’ *

Usually taken as ‘uncertainty on expt’

i) Makes algebra simpler

ii) If theory ~ expt, not too different

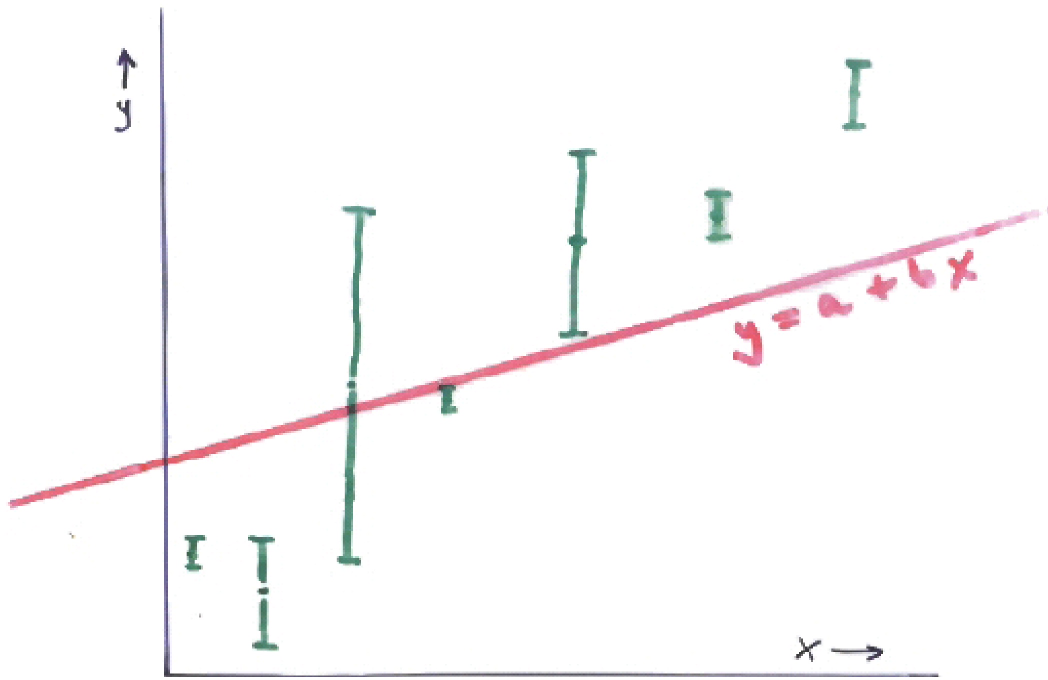
If theory and data OK:

$y^{th} \sim y^{obs} \rightarrow$ S small

Minimise S → best line

Value of S_{min} → how good fit is

* Th	Obs		σ_{th}	σ_{obs}	Contr to S
0.01	1	{	0.1		100
				1	1



$$\text{Criterion:} \quad S = \sum_i \left(\frac{\overbrace{y_i^{th}(a,b) - y_i^{obs}}^{\text{Vert devn}}}{\sigma_i} \right)^2$$

$a + bx_i$

An error for each pt.

Straight Line Fit

$$S = \sum_i \left(\frac{(a + b x_i) - y_i}{\sigma_i} \right)^2$$

i) "Draw" lots of lines $\Rightarrow$ S for each

ii) Minimise S (w.r.t. $\underline{\underline{a}}$ & $\underline{\underline{b}}$)

$$\frac{1}{2}\frac{\partial S}{\partial a} = \sum_i \frac{(a + b x_i - y_i)}{\sigma_i^2} = 0$$

$$\frac{1}{2}\frac{\partial S}{\partial b} = \sum_i \frac{(a + b x_i - y_i) x_i}{\sigma_i^2} = 0$$

2 SIM. EQNS FOR 2 UNKNOWNS ($\underline{\underline{a}}$ & $\underline{\underline{b}}$)

$$b = \frac{[1][xy] - [x][y]}{[1][x^2] - [x][x]} = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\langle x^2 \rangle - \langle x \rangle^2}$$

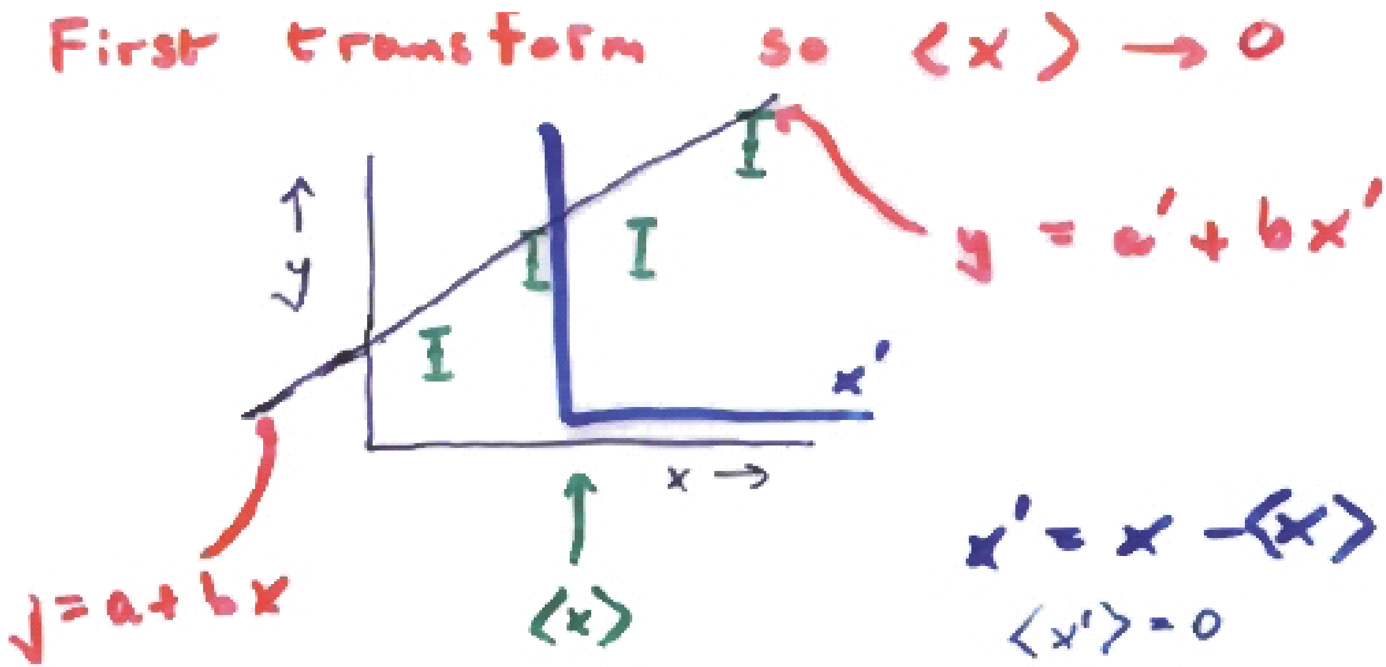
where $[f] = \sum \frac{f_i}{\sigma_i^2}$

& $\langle f \rangle = [f]/[1]$

$$\langle y \rangle = a + b \langle x \rangle$$

N.B. L.S.B.F. passes through (<x>, <y>)

Uncertainties on intercept and gradient



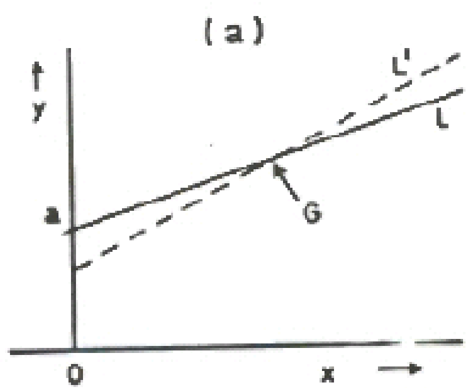
Better to use x' because

uncertainties on a' and b are UNCORRELATED

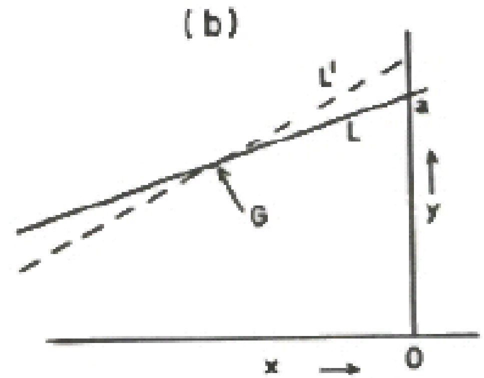
Contrast uncertainties on a and b are CORRELATED

That is why track parameters specified at track 'centre'

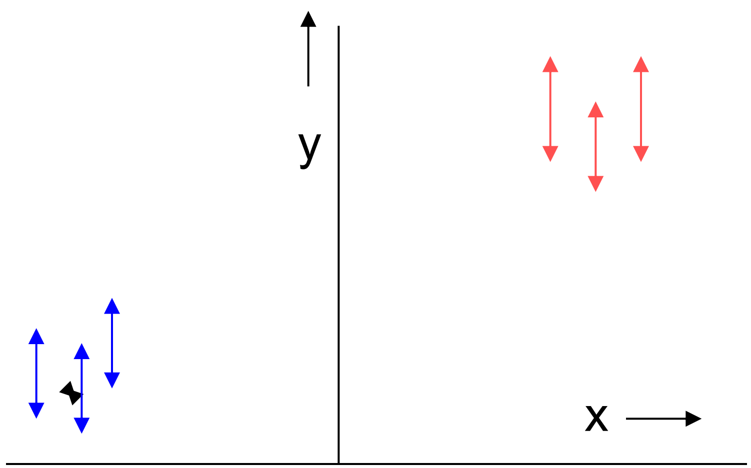
Covariance(a,b) ~ -<x>



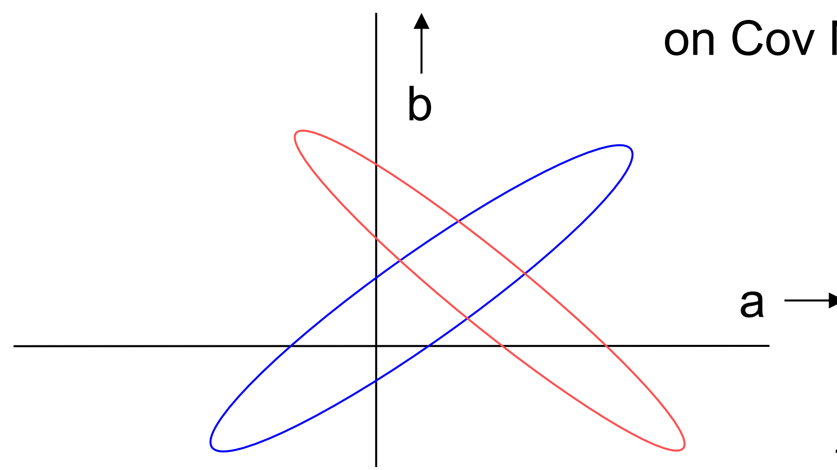
<x> positive



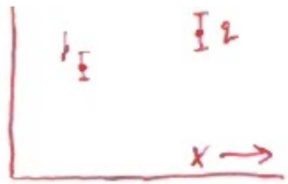
<x> negative



See lecture
on Cov Matrix



Measurements with correlated uncertainties **e.g. systematics**?



Start with 2 uncorrelated measurements

$$S = \frac{(p - p_{pr})^2}{\sigma_p^2} + \frac{(q - q_{pr})^2}{\sigma_q^2} \quad *$$

Introduce correlations by

$$\left.\begin{aligned} p &= r\cos\theta - s\sin\theta \\ q &= r\sin\theta + s\cos\theta \end{aligned}\right\}$$

NOT ROTN in x-y SPACE

Write σ_p σ_q $(+\ \mathrm{cov}(p,q) = 0)$ in terms of σ_r^2 σ_s^2 + $\mathrm{cov}(r,s)$

$$\Rightarrow S = \frac{1}{\sigma_r^2\sigma_s^2 - \mathrm{cov}(r,s)}\left[\sigma_s^2 (r - r_{pr})^2 + \sigma_r^2 (s - s_{pr})^2 - 2\,\mathrm{cov}(r,s)(r - r_{pr})(s - s_{pr})\right]$$

Inv. err matrix element →

$$= H_{11}(r - r_{pr})^2 + H_{22}(s - s_{pr})^2 + 2H_{12}(r - r_{pr})(s - s_{pr})$$

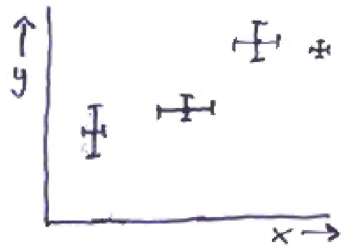
where $H^{-1} = \begin{pmatrix} \sigma_r^2 & \mathrm{cov} \\ \mathrm{cov} & \sigma_s^2 \end{pmatrix}$ ← Error matrix

Reduces to standard formula in absence of corrlns

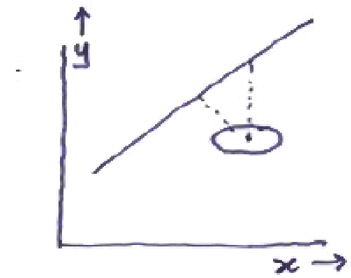
In general: $S = \sum_{ij} \tilde{\Delta}_i H_{ij} \Delta_j$

where $\Delta_j = (\text{observed} - \text{pred.})_j$

STRAIGHT LINE: Uncertainties on x and on y



For simplicity, assume x, y errors uncorrelated



Previously, contribution to S was $\left(\frac{y_i - y_i(fit)}{\sigma_i}\right)^2$

Now replace by

$$\text{Min}\left[\frac{\text{Distance of any point on line, to data point}}{\text{Radius of error ellipse in that dirn}}\right]^2$$

i.e. Min of error ellipse function

$$\frac{(x-x_i)^2}{\sigma_{x_i}^2} + \frac{(y-y_i)^2}{\sigma_{y_i}^2} = \frac{(y_i - a - b\,x_i)^2}{\sigma_{y_i}^2 + b^2\sigma_{x_i}^2}$$

Best line by minimising $S = \sum \frac{(y_i - a - b\,x_i)^2}{\sigma_{y_i}^2 + b^2\sigma_{x_i}^2}$

Errors as usual from $\frac{\partial^2 S}{\partial a^2}$ etc

Analytic soln if all σ_{x_i} same, & also σ_{y_i}

Comments on Least Squares method

1) Need to bin

Beware of too few events/bin (Want Poisson ~ Gaussian)

2) Extends to n dimensions →

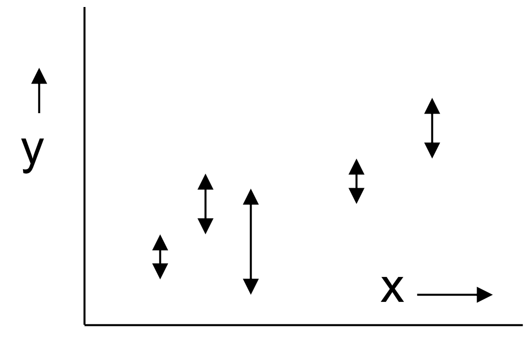
but needs lots of events for n larger than 2 or 3

3) No problem with correlated uncertainties

4) Can calculate S_{min} "on line" i.e. single pass through data

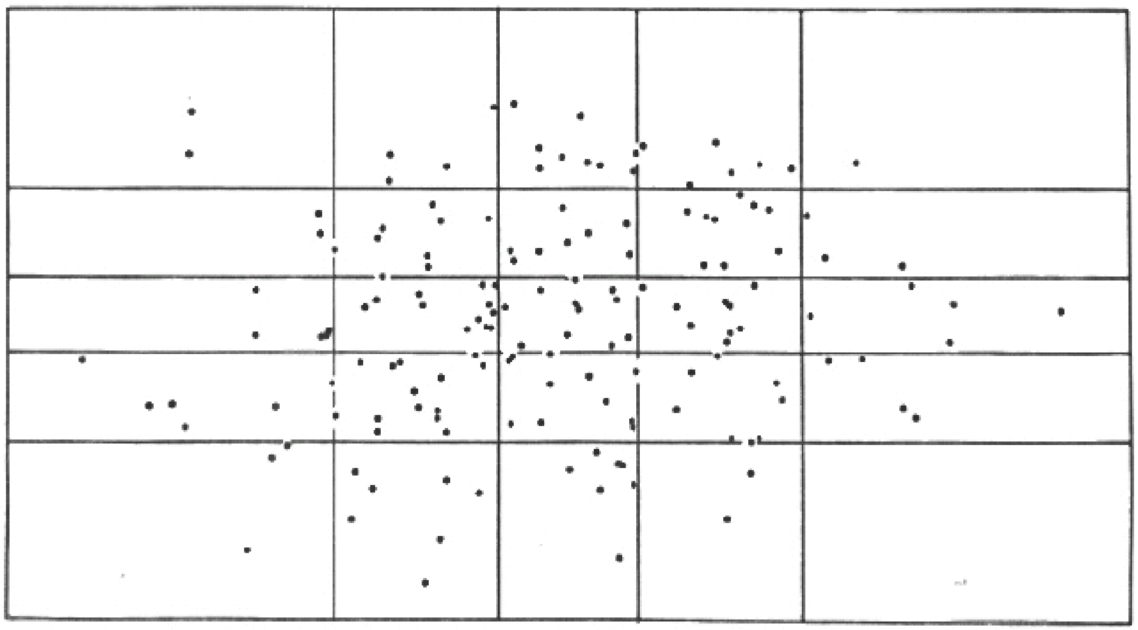
$$\Sigma (y_i - a - bx_i)^2 / \sigma^2 = [y_i^2] - b\,[x_i y_i] - a\,[y_i]$$

5) For theory linear in params, analytic solution



6) Goodness of Fit ★ ★ ★

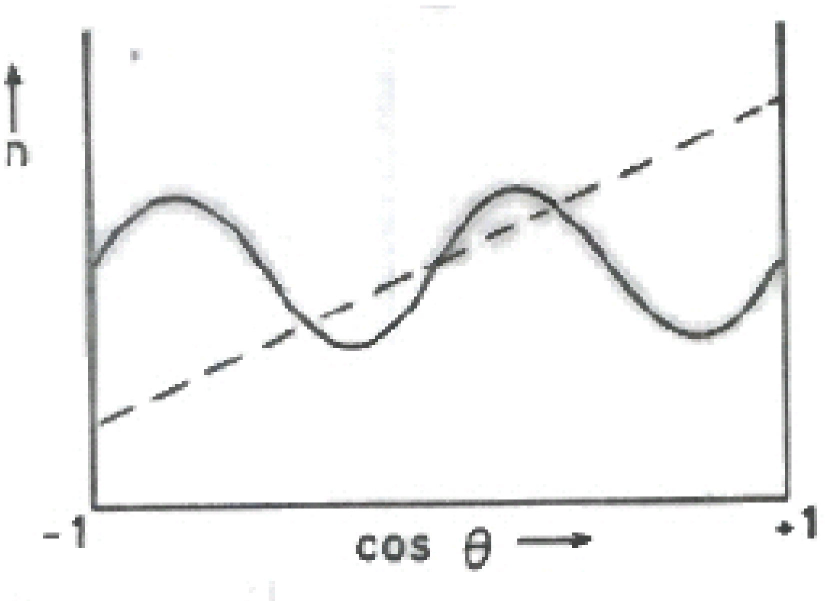
	Individual events (e.g. in cos θ)	$y_i \pm \sigma_i$ v x_i (e.g. stars)
1) Need to bin?	Yes	No need
4) χ^2 on line	First histogram	Yes



	Moments	Max Like	Least squares
Easy?	Yes, if…	Normalisation, maximisation messy	Minimisation
Efficient?	Not very	Usually best	Sometimes = Max Like
Input	Separate events	Separate events	Histogram
Goodness of fit	Messy	No (unbinned)	Easy
Constraints	No	Yes	Yes
N dimensions	Easy if ….	Norm, max messier	Easy
Weighted events	Easy	Errors difficult	Easy
Bgd subtraction	Easy	Troublesome	Easy
Uncertainty estimates	Observed spread, or analytic	$\left\{ - \frac{\partial^2 l}{\partial p_i \partial p_j} \right\}^{-1/2}$	$\left\{ \frac{\partial^2 S}{2 \partial p_i \partial p_j} \right\}^{-1/2}$
Main feature	Easy	Best	Goodness of Fit

‘Goodness of Fit’ by parameter testing?

$1+\beta\cos^2\theta$ Is $\beta = 0$?



‘Distribution testing’ is better

Goodness of Fit: χ^2 test

1) Construct S and minimise wrt free parameters
2) Determine ν = no. of degrees of freedom

$\nu = n - p$

n = no. of data points

p = no. of FREE parameters

3) Look up probability that, for ν degrees of freedom, $\chi^2 \geq S_{min}$

Works ASYMPTOTICALLY, otherwise use MC

[Assumes y_i are GAUSSIAN distributed with mean y_i^{th} and variance σ_i^2]

Properties of mathematical χ^2 distribution:

$$\overline{\chi^2} = \nu$$

$$\sigma^2(\chi^2) = 2\nu$$

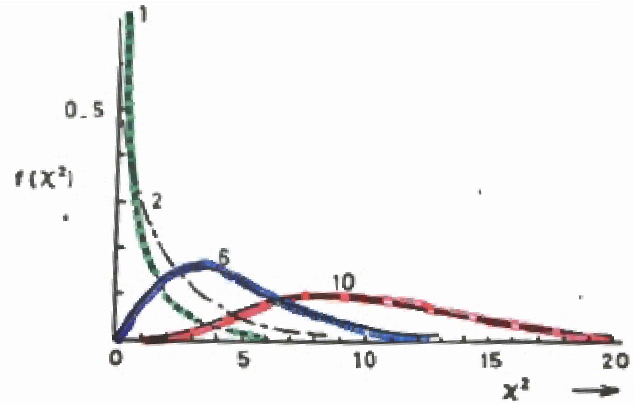
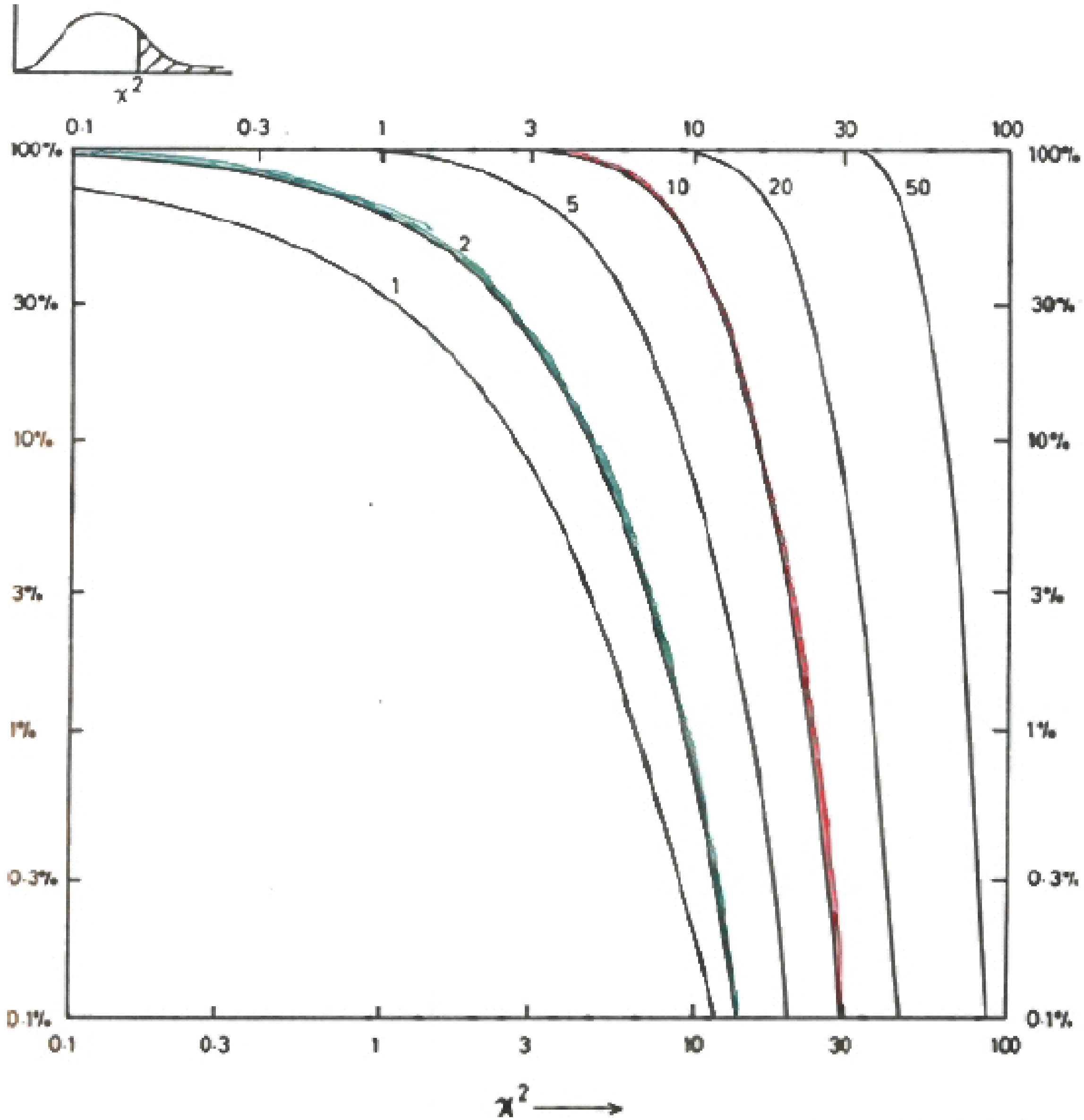


Fig. 2.6

So $S_{min} > \nu + 3\sqrt{2\nu}$ is LARGE

e.g. $S_{min} = 2200$ for $\nu = 2000$?



Cf: Area in tails of Gaussian

χ^2 with ν degrees of freedom?

ν = data – free parameters ?

Why asymptotic (apart from Poisson → Gaussian) ?

a) Fit flatish histogram with

$y = N\ \{1 + 10^{-6} \cos(x - x_0)\}$ x_0 = free param

b) Neutrino oscillations: almost degenerate parameters

$y \sim 1 - A \sin^2(1.27\ \Delta m^2\ L/E)$ 2 parameters

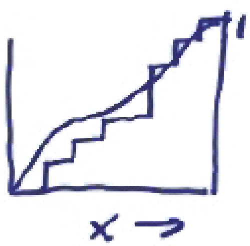
$\xrightarrow[\text{Small } \Delta m^2]{} 1 - A\ (1.27\ \Delta m^2\ L/E)^2$ 1 parameter

Goodness of Fit

. $\chi 2$ Very general
Needs binning
Not sensitive to sign of deviation

Run Test

Kolmogorov-Smirnov

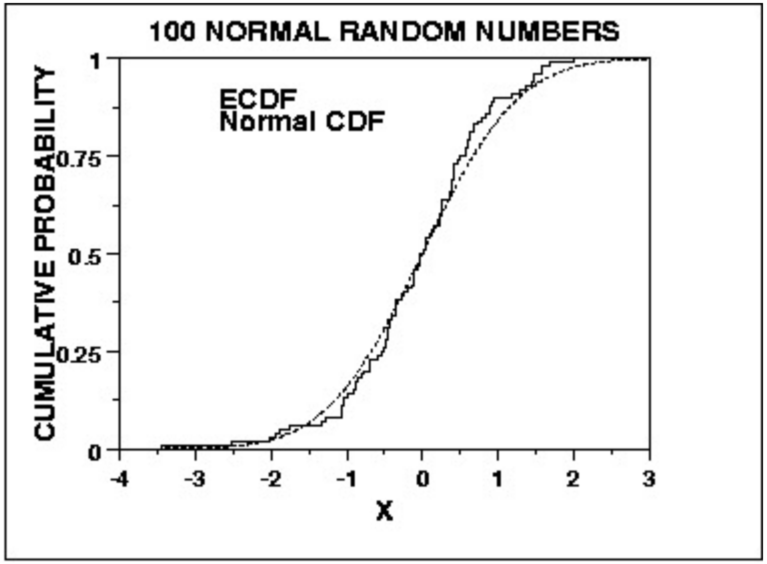


Aslan and Zech
Durham IPPP Stats Conf

etc

Goodness of Fit: Kolmogorov-Smirnov

Compares data and model cumulative plots
(or 2 sets of data)
Uses largest discrepancy between dists.
Model can be analytic or MC sample



Uses individual data points
Not so sensitive to deviations in tails
(so variants of K-S exist)
Not readily extendible to more dimensions
Distribution-free conversion to p; depends on n
(but not when free parameters involved – needs MC)

Goodness of fit: ‘Energy’ test

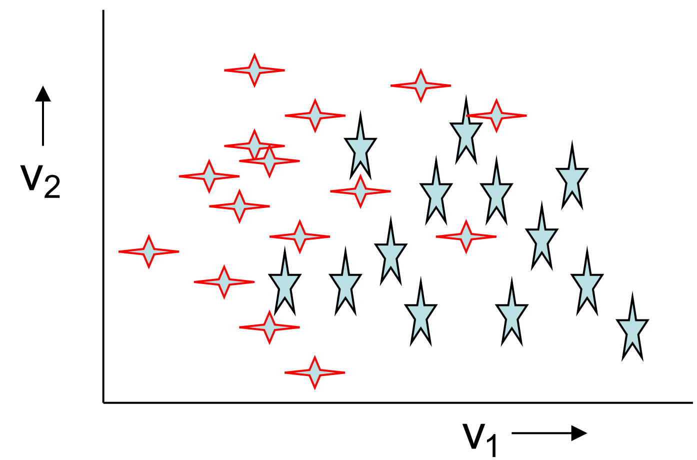
Assign +ve charge to data ; -ve charge to M.C.

Calculate ‘electrostatic energy E’ of charges

If distributions agree, E ~ 0

If distributions don’t overlap, E is positive

Assess significance of magnitude of E by MC



N.B.

1) Works in many dimensions

2) Needs metric for each variable (make variances similar?)

3) $E \sim \Sigma\, q_i q_j\, f(\Delta r = |r_i - r_j|)$, $f = 1/(\Delta r + \varepsilon)$ or $-\ln(\Delta r + \varepsilon)$

   Performance insensitive to choice of small ε

See Aslan and Zech’s paper at:
  http://www.ippp.dur.ac.uk/Workshops/02/statistics/program.shtml

Wrong Decisions

Error of First Kind

Reject H0 when true

Should happen x% of tests

Errors of Second Kind

Accept H0 when something else is true

Frequency depends on ………

i) How similar other hypotheses are

e.g. H0 = μ

Alternatives are:	e		π	K	p
ii) Relative frequencies:	10^{-4}	10^{-4}	1	0.1	0.1

Aim for maximum efficiency ⟵ Low error of 1st kind

maximum purity ⟵ Low error of 2nd kind

As χ^2 cut tightens, efficiency ↑ and purity ↓

Choose compromise

How serious are errors of 1st and 2nd kind?

1) Result of experiment

e.g Is spin of resonance = 2?

Get answer WRONG

Where to set cut?

Small cut ⇒ Reject when correct

Large cut ⇒ Never reject anything

Depends on nature of H0 e.g.

Does answer agree with previous expt?

Is expt consistent with special relativity?

2) Class selector e.g. b-quark / galaxy type / γ-induced cosmic shower

Error of 1st kind: Loss of efficiency

Error of 2nd kind: More background

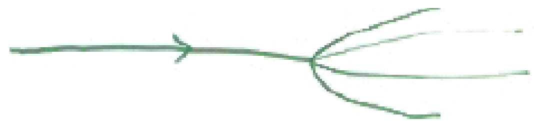
Usually easier to allow for 1st than for 2nd

3) Track finding

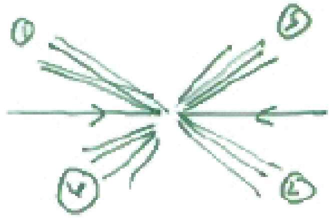
KINEMATIC FITTING

Tests whether observed event is consistent with specified reaction

e.g.



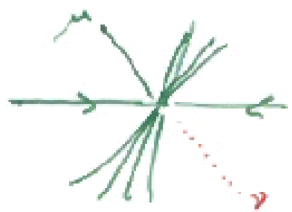
$\bar{p}p \to \bar{p}p\pi^+\pi^-$?



$e^+e^- \to W^+W^-$

$\hookrightarrow j_1 j_2 \qquad \to j_3 j_4$

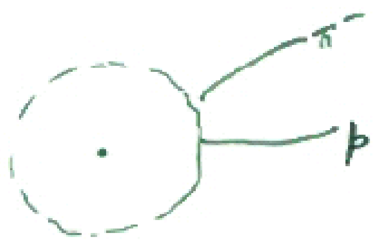
M_W, jet pairings



$e^+e^- \to W^+W^-$

$\hookrightarrow j_1 j_2 \qquad \to \mu\nu$

$\Lambda \to p\pi^-$ from prodn vertex



$p + \pi^-$ intersect

& $\Lambda \to p\pi^-$ from prodn vert.

Kinematic Fitting: Why do it?

1) Check whether event consistent with hypothesis [Goodness of Fit]

2) Can calculate missing quantities [Param detn.]

3) Good to have tracks conserving E-P [Param detn.]

4) Reduces uncertainties [Param detn.]

Kinematic Fitting: Why do it?

1) Check whether event consistent with hypothesis [Goodness of Fit]
Use S_{min} and ndf

2) Can calculate missing quantities [Param detn.]
e.g. Can obtain |P| for short/straight track, neutral beam; p_x,p_y,p_z of outgoing ν, n, K^0

3) Good to have tracks conserving E-P [Param detn.]
e.g. identical values for resonance mass from prodn or decay

4) Reduces uncertainties [Param detn.]
Example of “Including theoretical input reduces uncertainties”

How we perform Kinematic Fitting ?

Observed event: 4 outgoing charged tracks
Assumed reaction: $pp \rightarrow pp\pi^+\pi^-$

Measured variables: 4-momenta of each track, v_i^{meas}
(i.e. 3-momenta & assumed mass)
Then test hypothesis:
Observed event = example of assumed reaction

i.e. Can tracks be wiggled "a bit" to do so?

Tested by:

$$S_{min} = \Sigma(v_i^{fitted} - v_i^{meas})^2 / \sigma^2$$

where v_i^{fitted} conserve 4-momenta
(Σ over 4 components of each track)
N.B. Really need to take correlations into account

i.e. Minimisation subject to constraints (involves Lagrange Multipliers)

Toy example of Kinematic Fit

$\bar{p}p \to \bar{p}p$

Fixed target experiment

+ constraints:

1) Coplanar
2) p_1 at θ_1
3) p_2 at θ_2
4) θ_1 at θ_2 $\Longleftarrow$ Non-relativistic equal mass elastic scatter : $\theta_1 + \theta_2 = \pi/2$

Measured	$\theta_1^m \pm \sigma$	$\theta_2^m \pm \sigma$
Fitted	θ_1	θ_2

Minimise $S(\theta_1, \theta_2) = \dfrac{(\theta_1 - \theta_1^m)^2}{\sigma^2} + \dfrac{(\theta_2 - \theta_2^m)^2}{\sigma^2}$

subject to $C(\theta_1, \theta_2) = \theta_1 + \theta_2 - \pi/2 = 0$

Lagrange : $\dfrac{\partial S}{\partial \theta_1} + \lambda \dfrac{\partial C}{\partial \theta_1} = \dfrac{\partial S}{\partial \theta_2} + \lambda \dfrac{\partial C}{\partial \theta_2} = 0$

$\Rightarrow$ 3 eqns for θ_1 θ_2 λ

Eqns simple to solve because

$c(\theta_1, \theta_2)$ linear in θ_1, θ_2

$$\Rightarrow \theta_1 = \theta_1^m + \tfrac{1}{2}(\pi/2 - \theta_1^m - \theta_2^m)$$

$$\theta_2 = \theta_2^m + \tfrac{1}{2}(\pi/2 - \theta_1^m - \theta_2^m)$$

$$\sigma(\theta_1) = \sigma(\theta_2) = \sigma/\sqrt{2}$$

i.e. KINEMATIC FIT → REDUCED UNCERTAINTIES

‘KINEMATIC’ FITTING

Angles of triangle: $\theta_1 + \theta_2 + \theta_3 = 180$

	θ_1	θ_2	θ_3	
Measured	50	60	73 ± 1	Sum = 183
Fitted	49	59	72	180

$\chi^2 = (50\text{-}49)^2/1^2 + 1 + 1 = 3$

Prob $\{\chi^2_1 > 3\} = 8.3\%$

ALTERNATIVELY:

Sum $= 183 \pm 1.7$, while expect 180

Prob{Gaussian 2-tail area beyond 1.73σ} = 8.3%

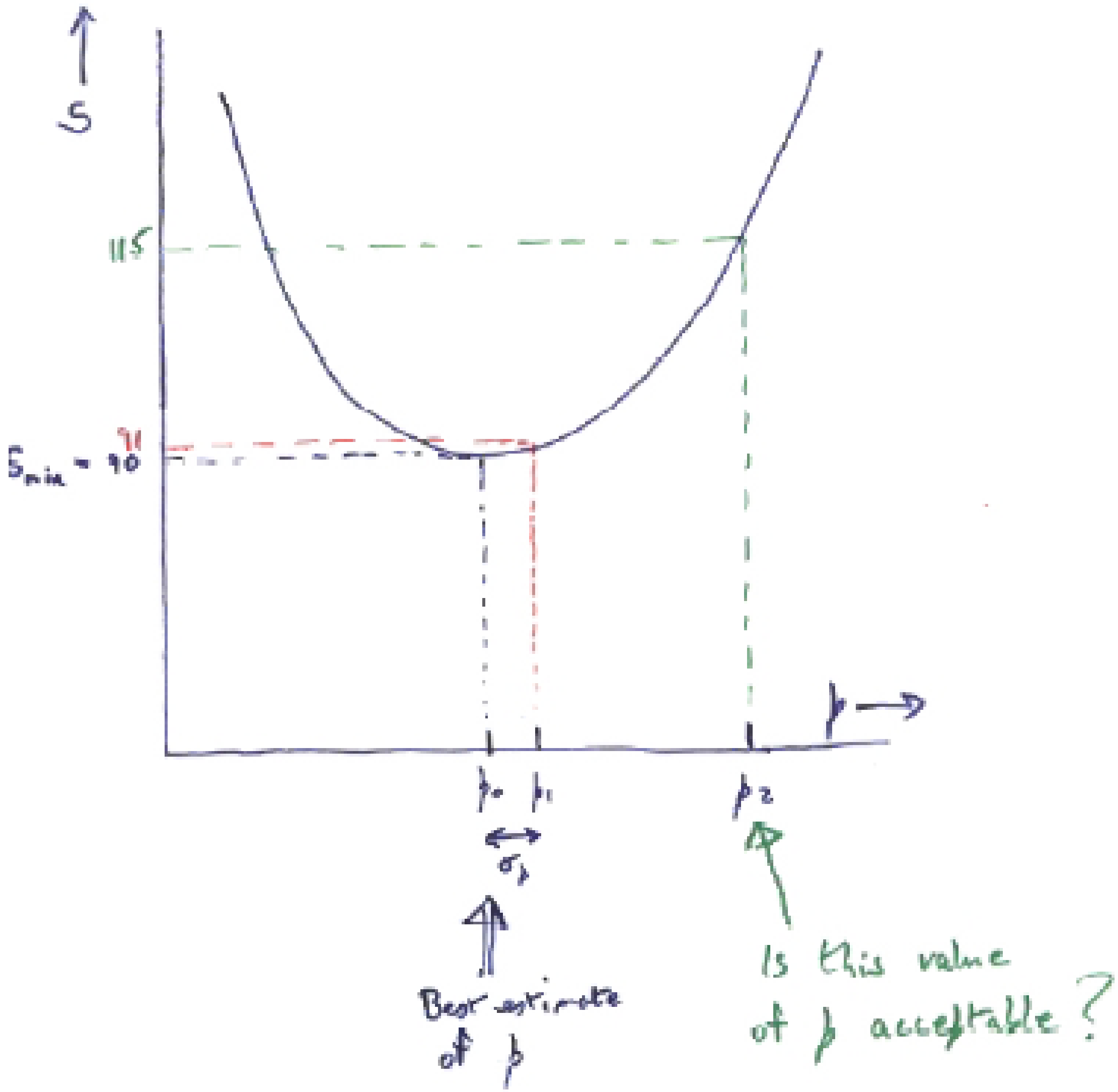
THE PARADOX

Histogram with 100 bins
Fit with 1 parameter
S_{min}: χ^2 with NDF = 99 (Expected $\chi^2 = 99 \pm 14$)

For our data, $S_{min}(p_0) = 90$
Is p_2 acceptable if $S(p_2) = 115$?

1) YES. Very acceptable χ^2 probability

2) NO. σ_p from $S(p_0 + \sigma_p) = S_{min} + 1 = 91$
But $S(p_2) - S(p_0) = 25$
So p_2 is 5σ away from best value


S
115
91
S_min = 90
p
p_0
p_1
p_2
σ_p
Best estimate of p
Is this value of p acceptable?