FNAL STATISTICS LECTURE #3

LEAST SQUARES BEST FIT

- STRAIGHT LINE
- CORRELATED ERRORS
- ERRORS IN X AND Y

HYPOTHESIS TESTING BY χ^2

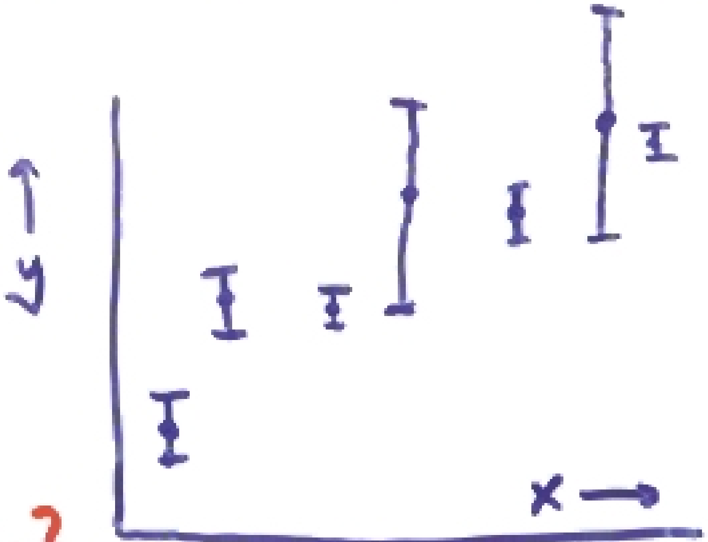
- ERRORS OF FIRST + SECOND KIND
- KINEMATIC FITTING
    - TOY EXAMPLE

THE PARADOX

LOUIS LYONS

AUG 2004

LEAST SQUARES STRAIGHT LINE FITTING



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

Th: $y = a + bx$

1) DOES IT FIT STRAIGHT LINE?

(HYPOTHESIS TESTING)

2) WHAT ARE GRADIENT + INTERCEPT?

(PARAMETER DETERMINATION

1st

N.B. 1 CAN BE USED FOR NON - "$a + bx$"

e.g. $a + b\cos^2\theta$

N.B. 2. LEAST SQUARES NOT ONLY METHOD

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

σ_i SUPPOSED TO BE "ERROR ON TH." *

TAKEN AS "ERROR ON EXPT"

i) Makes algebra simpler

ii) If theory ~ expt, not too different.

IF THEORY (& DATA) O.K.

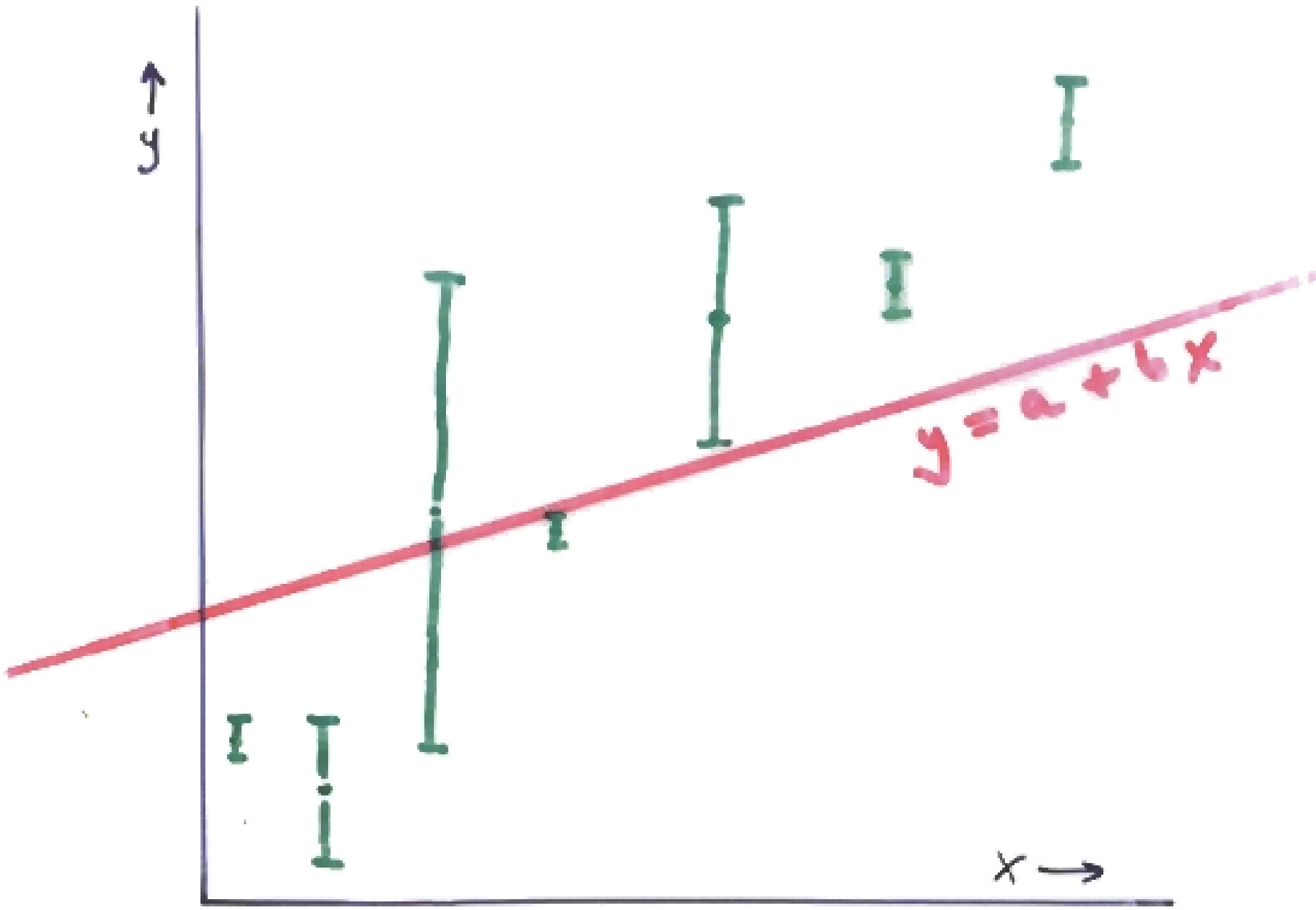
$y^{th} \sim y^{obs} \Longrightarrow S$ small

Minimise $S \Rightarrow$ best line

Value of $S_{min} \Rightarrow$ how good fit is.

*

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



$a + bx_i$

Criterion:

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

An error for each pt.

SIMPLE EXAMPLE OF MINIMISING S

Measurements $a_1 \pm \sigma_1$, $a_2 \pm \sigma_2$, $\vdots$, $a_i \pm \sigma_i$ $\Big\}$ Best value $\hat{a} \pm \sigma$

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

Minimise S w.r.t. $\hat{a}$

$$\frac{1}{2}\frac{\partial S}{\partial \hat{a}} = \sum \frac{\hat{a} - a_i}{\sigma_i^2} = 0$$

$$\hat{a} \sum \frac{1}{\sigma_i^2} = \sum \frac{a_i}{\sigma_i^2}$$

Error on $\hat{a}$ given by $\sigma = \left(\frac{1}{2}\frac{\partial^2 S}{\partial \hat{a}^2}\right)^{-1/2}$

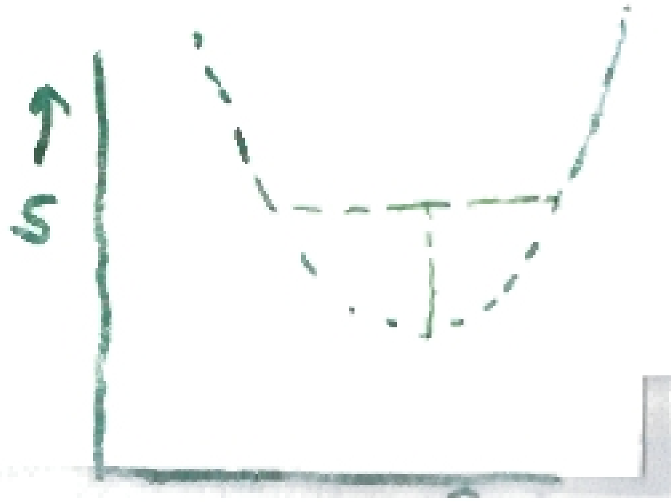
IN PARABOLIC APP
EQUIV TO
$S \to S_{min} + 1$

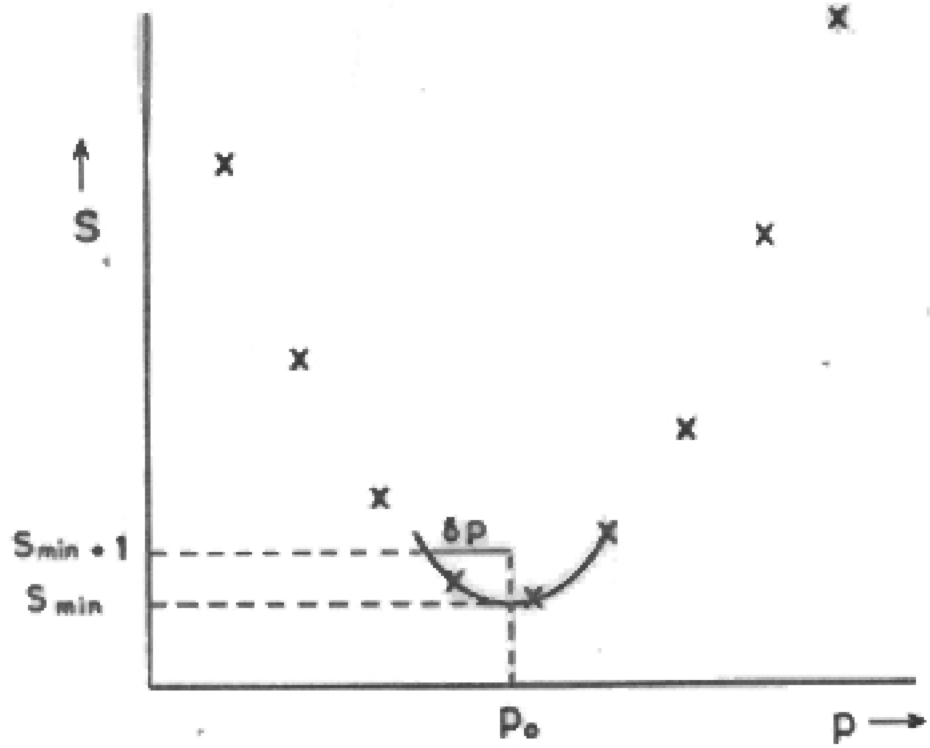
$$\frac{\partial^2 S}{\partial \hat{a}^2} = 2 \sum \frac{1}{\sigma_i^2}$$

$$\therefore \frac{1}{\sigma^2} = \sum \frac{1}{\sigma_i^2}$$

Many params

$$\frac{1}{2}\frac{\partial^2 S}{\partial p_i \partial p_j} = \text{INVERSE ERROR MATRIX}$$




S
Smin + 1
Smin
δp
po
p

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

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

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

$$\left.\begin{aligned} \frac{1}{2}\frac{\partial S}{\partial a} &= \sum_i \frac{(a + bx_i - y_i)}{\sigma_i^2} = 0 \\ \frac{1}{2}\frac{\partial S}{\partial b} &= \sum_i \frac{(a + bx_i - y_i)\,x_i}{\sigma_i^2} = 0 \end{aligned}\right\}$$

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

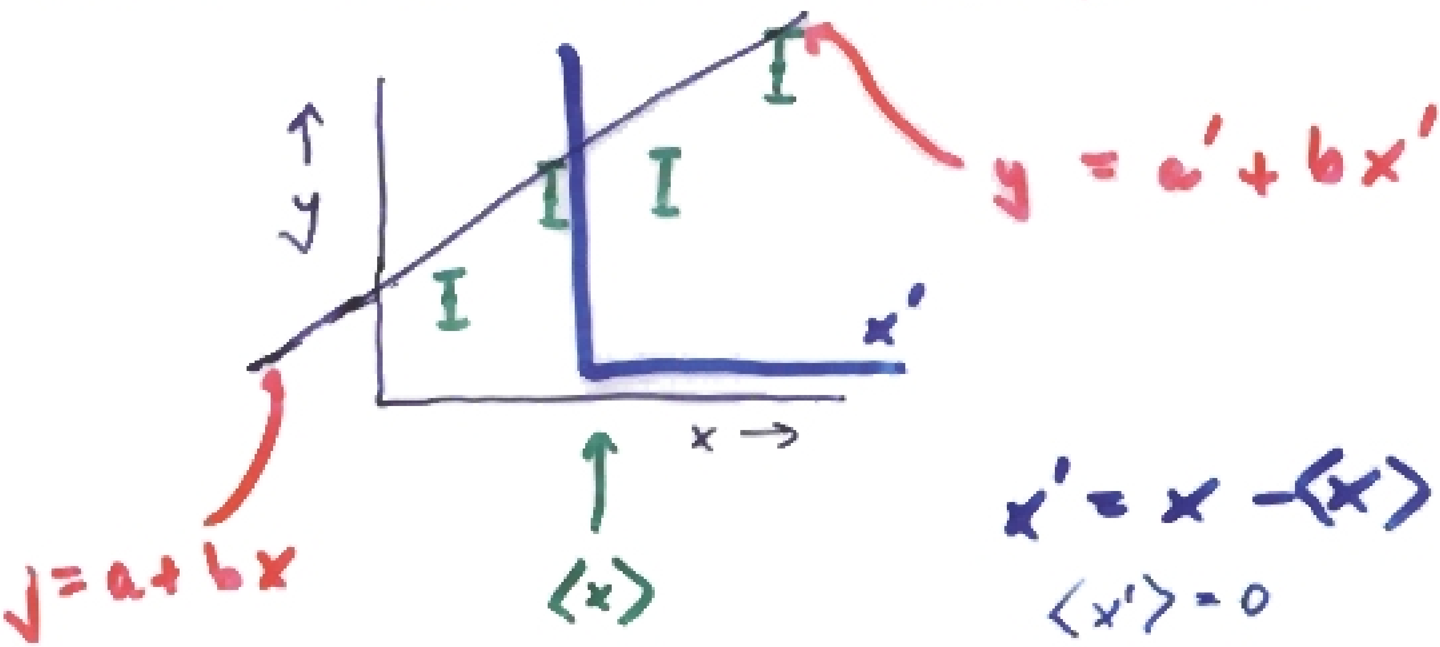
& $\langle f \rangle = [f]/[1]$

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

N.B. L.S.B.F. line passes through $(\langle x \rangle, \langle y \rangle)$

Error on intercept & gradient

First transform so $\langle x \rangle \to 0$



$x' = x - \langle x \rangle$

$\langle x' \rangle = 0$

Better to use x' because

error on a' & b are UNCORRELATED

[cf. Errors on a & b CORRELATED]

$$\left.\begin{aligned} \sigma(a') &= 1\Big/\sqrt{\tfrac{1}{2}\frac{\partial^2 S}{\partial a'^2}} \\ \sigma(b) &= 1\Big/\sqrt{\tfrac{1}{2}\frac{\partial^2 S}{\partial b^2}} \end{aligned}\right\} \Leftarrow \operatorname{cov}(a', b) = 0$$

$$S = \sum_i \left(\frac{a' + b x_i' - y_i}{\sigma_i} \right)^2 = a'^2[1] + b^2[x'^2]$$

$+ [y^2]$ + cross-term (inc $a'b[x']$)

$$\left.\begin{aligned} \sigma^2(a') &= 1/[1] \\ \sigma^2(b) &= 1/[x'^2] \end{aligned}\right\}$$

N.B. Errors depend on σ_i, but **NOT** on how well data agrees with theory

For error on y at other x', use $y = a' + bx'$

$$\Rightarrow \sigma^2(y) = \sigma^2(a') + x'^2 \sigma^2(b)$$

For $x' = -\langle x \rangle$ [i.e. $x = 0$]

$$\sigma^2(a) = \sigma^2(a') + \langle x \rangle^2 \sigma^2(b)$$

BUT $\sigma(a)$ + $\sigma(b)$ CORRELATED

SPECIAL CASE : All σ_i same

$$\sigma^2(a') = \sigma^2/n$$

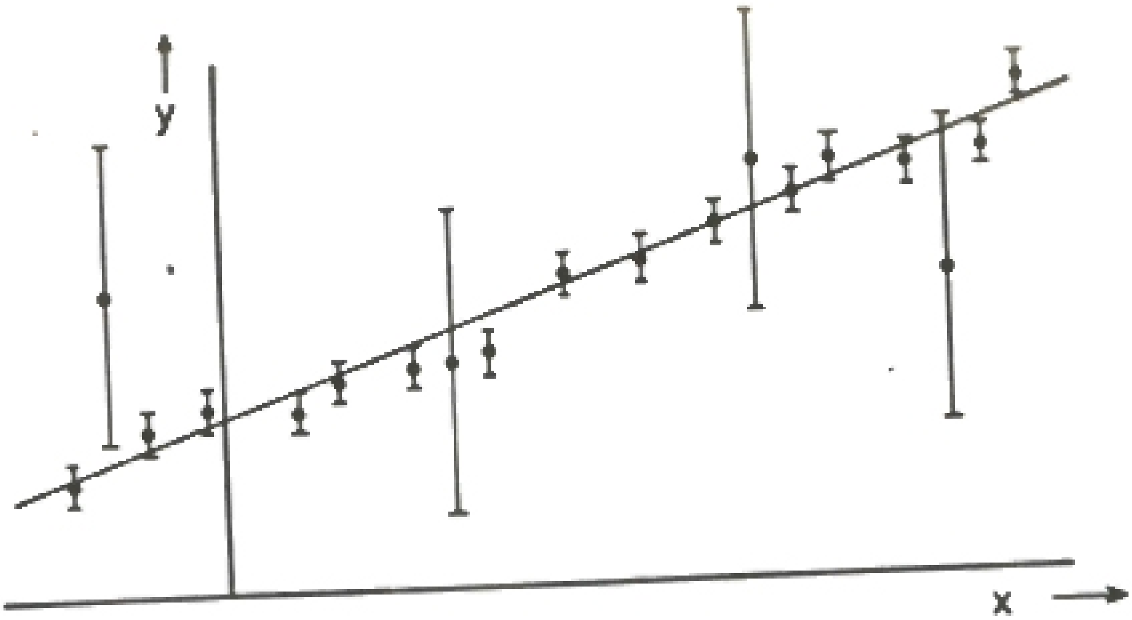
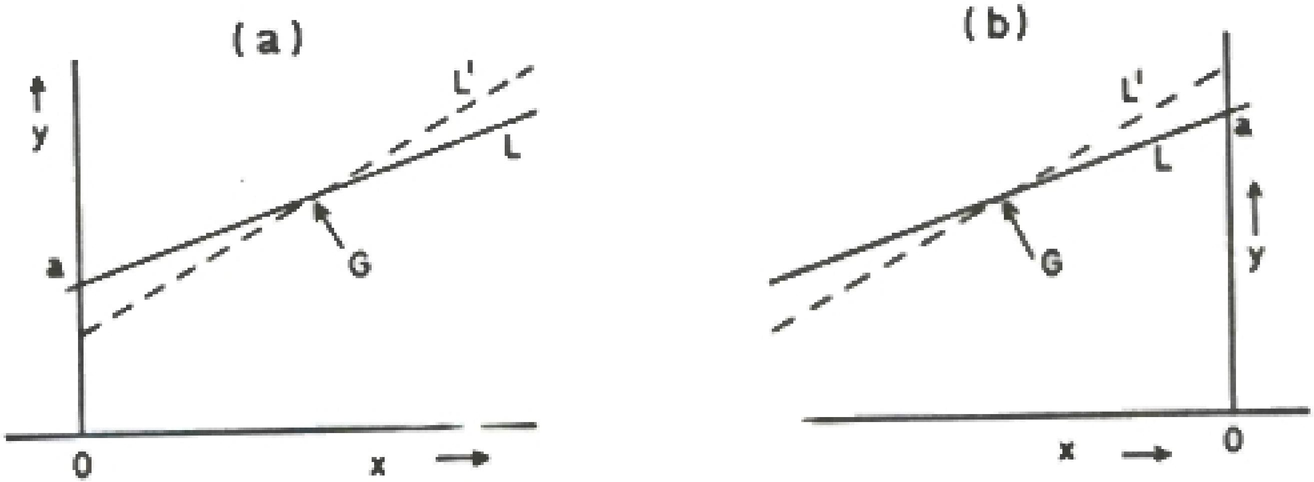


Fig. 2.3

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



$\langle x \rangle$ pos

Fig. 2.4

$\langle x \rangle$ neg

IF NO ERRORS δy_i (!)

ASSUME ALL ERRORS EQUAL

(or similar)

σ CANCELS FROM a + b

e.g. $$b = \frac{[1][xy] - [x][y]}{[1][x^2] - [x]^2}$$

NEED σ for errors on a' + b

$$S = \frac{1}{\sigma^2} \sum (a + bx_i - y_i)^2 = \nu$$

$$\Rightarrow \sigma$$

$$\Rightarrow \sigma(a') + \sigma(b)$$

i.e USE SCATTER OF POINTS AROUND STRAIGHT LINE $\Rightarrow$ ERROR ON POINTS

$\Rightarrow$ ERROR ON INTERCEPT + GRADIENT

(cf: Estimate σ from scatter of repeated measurements)

N.B. CANNOT TEST WHETHER DATA IS CONSISTENT WITH THEORY

SUMMARY OF STRAIGHT LINE FIT

1) PLOT DATA
   a) BAD POINTS
   b) a AND b, & $\sigma(a')$, $\sigma(b)$

2) a AND b FROM FORMULAE*

3) ERRORS ON a' AND b *

4) CF 2) and 3) WITH 1)

5) DETERMINE S_{min} (using a & b)*

6) $\nu = n - p$ *

7) Look up χ^2 tables*

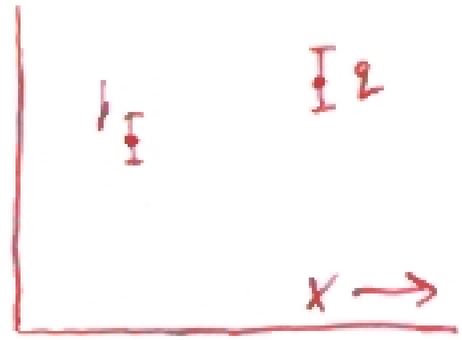
8) IF PROBABILITY TOO SMALL, IGNORE RESULTS

8a) IF PROBABILITY IS "A BIT" SMALL, SCALE ERRORS?

* COMPUTER PROGRAMME

MEASUREMENTS WITH CORRELATED ERRORS

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

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

Inv. err matrix element $\longrightarrow$

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

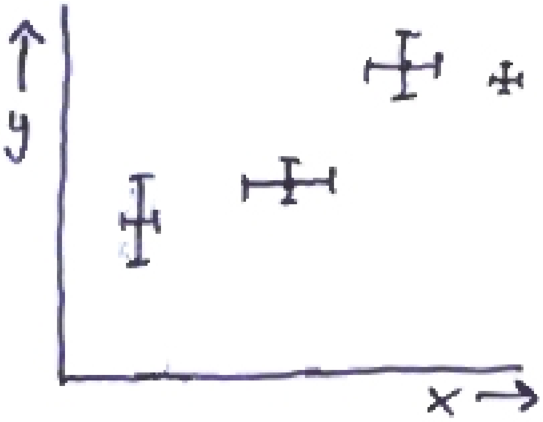
where $H^{-1} = \begin{pmatrix} \sigma_r^2 & \operatorname{cov} \\ \operatorname{cov} & \sigma_s^2 \end{pmatrix} \leftarrow$ error matrix

Reduces to standard formula in absence of corrlns

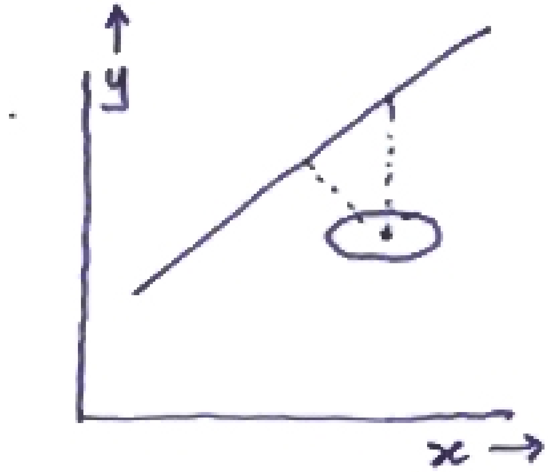
In general : $$S = \sum_{ij} \tilde{\Delta}_i H_{ij} \Delta_j$$

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

STRAIGHT LINE : ERRORS ON X AND Y



For simplicity, assume x, y errors uncorrelated



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

Now replace by

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

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

2) Extends to n dimensions $\Rightarrow$

   but needs lots of events for $n \gtrsim 3$

3) No problem with correlated errors

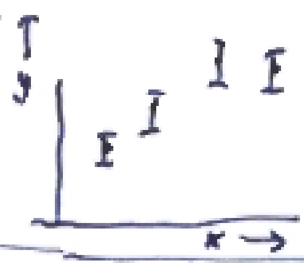
4) Can calculate χ^2 "on line" (i.e. single pass through data)

$$\sum \frac{(y_i - a - bx)^2}{\sigma^2} = [y_i^2] - b[x_i y_i] - a[y_i]$$

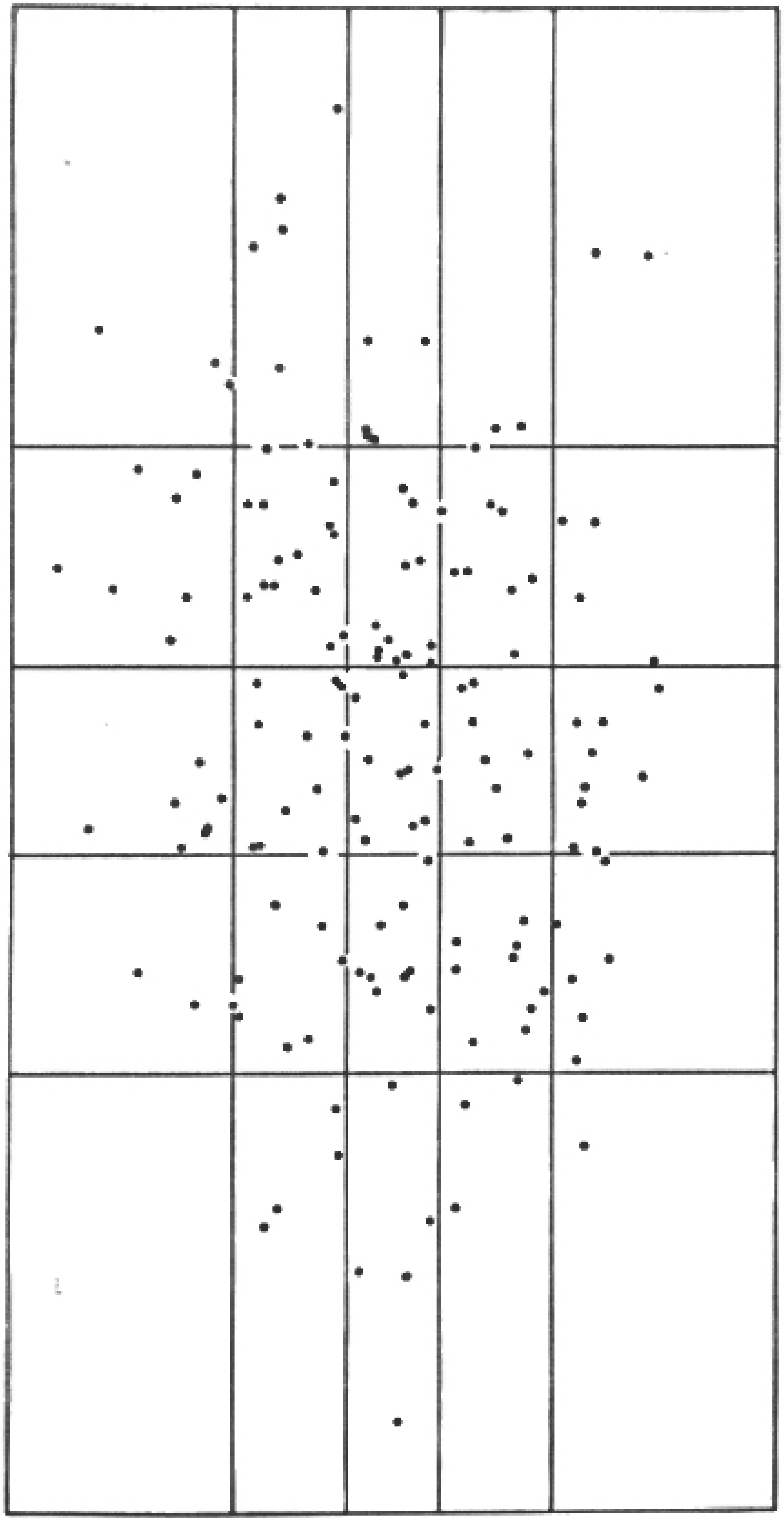
5) For theory linear in parameters,

   soln can be found analytically

6) Hypothesis Testing ✡ ✡ ✡



	Individual events (e.g in $\cos\theta$)	$y_i \pm \sigma_i$ v x_i (e.g. stars)
1) Binning first	✓	✗
4) χ^2 on line	First histogram	✓

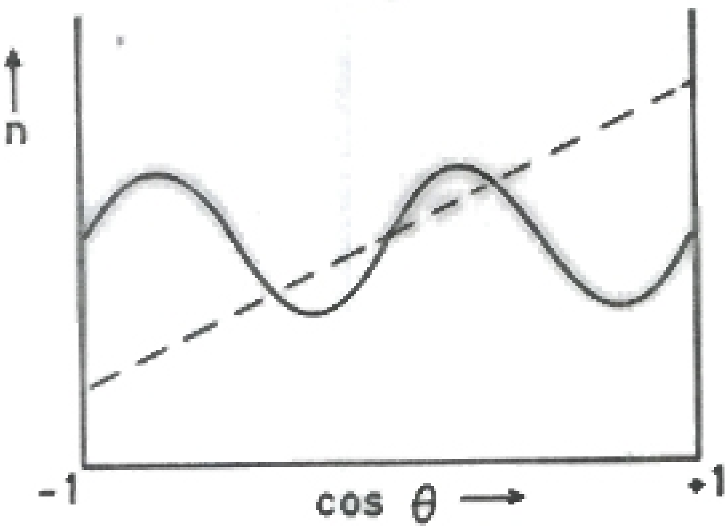


	Mom.	M.L.	L.S.
Easy?	Yes, if...	Norm, norm. messy	Minimisation
Efficient?	Not very	Usually best	Sometimes $\equiv$ M.L.
Input	Separate events.	Separate ev.	Histogram
Goodness of Fit	Messy	V. difficult	Easy
Constraints	No	Easy	Can be done
n-dimensions	Easy, if...	Norm, norm messier	Needs v. many events
Weighted ev.	Easy	Errors diff.	Easy
Bgd sub	Easy	Troublesome	Easy
Error est.	Observed spread OR Analytic	$\left(-\frac{\partial^2 \ell}{\partial p_i \partial p_j}\right)^{-\frac{1}{2}}$	$\left(\frac{1}{2}\frac{\partial^2 S}{\partial p_i \partial p_j}\right)^{-\frac{1}{2}}$
Main +	EASY	BEST FEW EVENTS	HYP. TEST.

HYPOTHESIS TESTING

by PARAMETER TESTING

$$1 + \frac{b}{a}\cos^2\theta \qquad \text{Is } \frac{b}{a} = 0?$$



"DISTRIBUTION TESTING" IS BETTER

HYPOTHESIS TESTING

χ^2 TEST

1) CONSTRUCT S, + MINIMISE W.R.T. FREE PARAMETERS

2) DETERMINE ν = NO. OF DEGREES OF FREEDOM

$$\nu = n - p$$

n = NO OF DATA POINTS

p = NO OF FREE PARAMS

3) LOOK UP PROB THAT, FOR ν DEG OF FREEDOM, $\chi^2 \geq S_{min}$

[ASSUMES y_i ARE GAUSSIAN DISTRIBUTED WITH MEAN y_i^{th} AND VARIANCE σ_i^2]

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

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

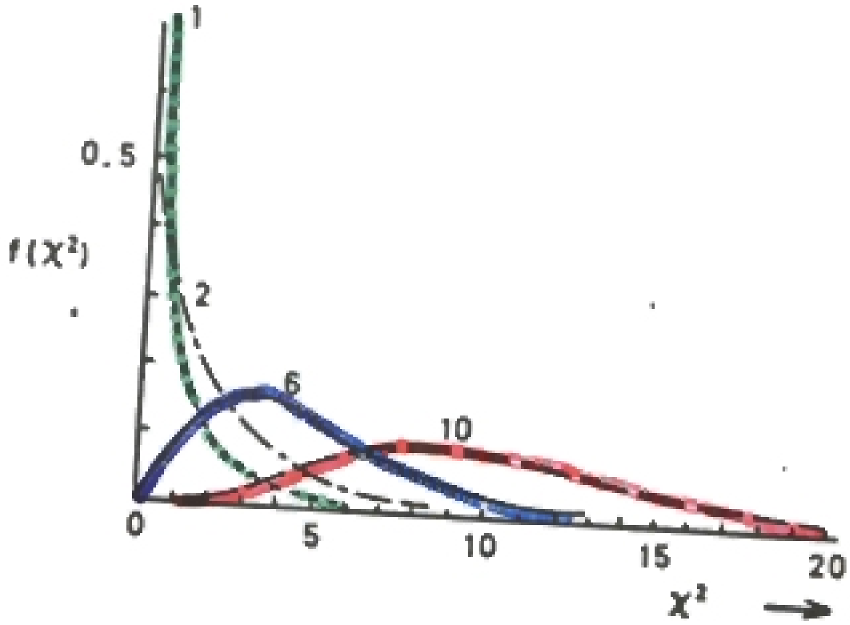


Fig. 2.6

$$\therefore \; S_{min} \gtrsim \nu + 3\sqrt{2\nu}$$

is <u>LARGE</u>

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

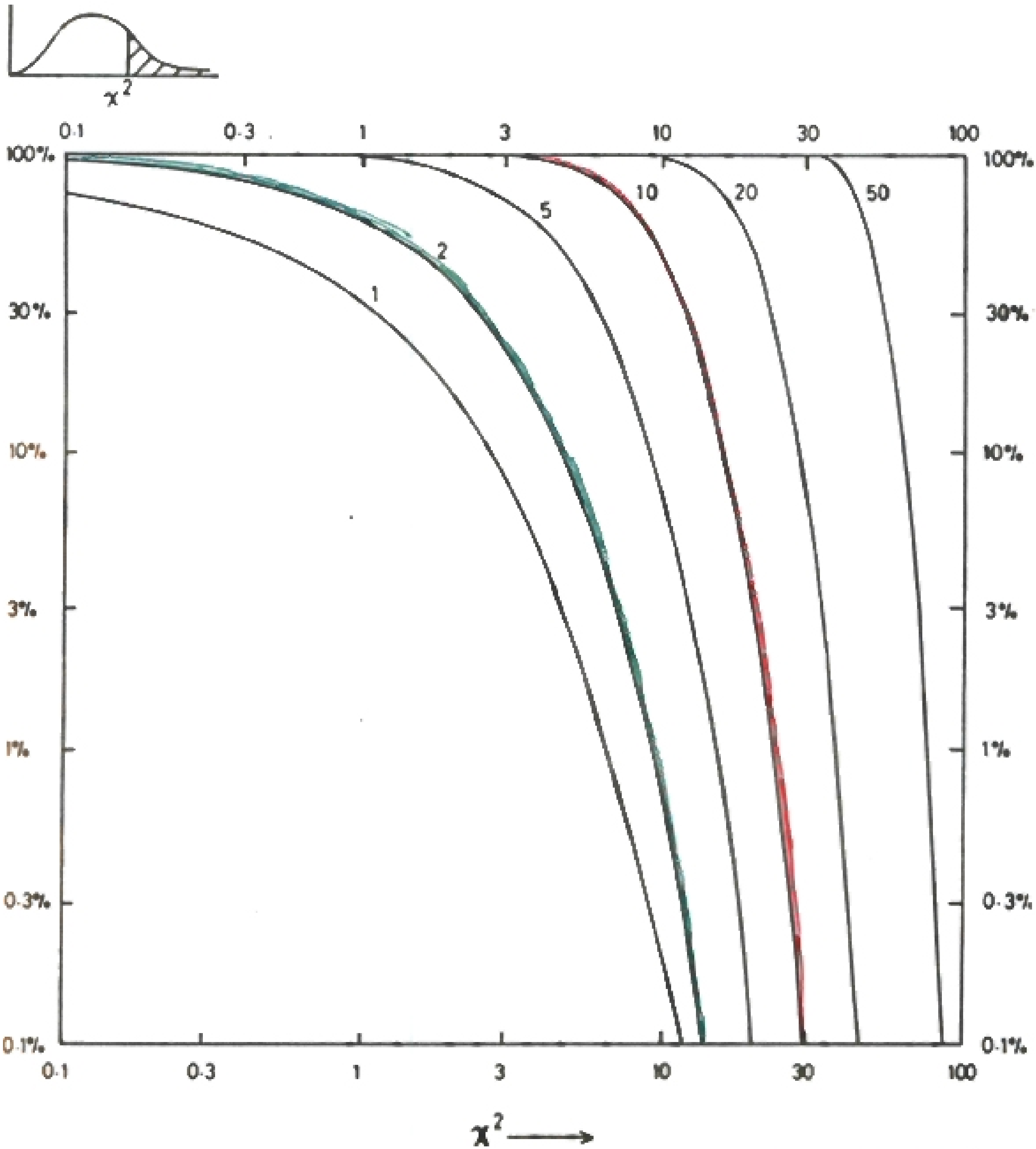


Fig. 2.7

CF: Area in tails of Gaussian

Goodness of Fit

χ^2 : Very general

Needs binning

Not sensitive to sign of devn.

Run test

Kolmogorov - Smirnov

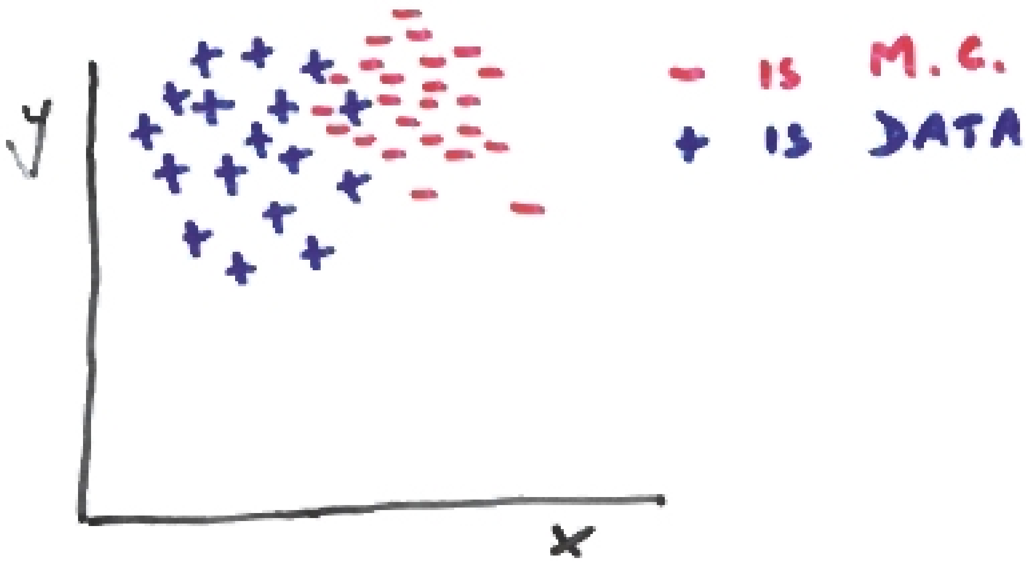
etc

See: Aslan + Zech, Durham IPPP Statistics Conf (2002)

Maria Grazia Pia's group in Genoa

ENERGY TEST FOR GOODNESS OF FIT

Aslan + Zech



$$\text{"Energy"} = \sum_{+} q_i \, q_j \, f(|r_{ij}|)$$

$$f = \frac{1}{r + \varepsilon}$$

$$\text{or } -\ln(r + \epsilon)$$

N.B. ϵ, choice of f

Scaling of $x, y, \ldots$

Need M.C.

WRONG DECISIONS

ERROR OF FIRST KIND

Reject H when it is true

Should happen x% of time

ERROR OF SECOND KIND

Accept H when something else is true

How often depends on .....

i) How similar other hypotheses are

e.g. $H = \pi$

Alternatives $= e \quad \mu \quad K \quad p \ldots$

ii) Relative frequencies

e.g. $10^{-4} \quad 10^{-4} \quad 10\% \quad 10\%$

Aim for maximum effic ← small error 1^{st} kind

maximum purity ← small error 2^{nd} kind

As χ^2 cut increases, effic ↑ purity ↓

Choose compromise

HOW SERIOUS ARE ERRORS OF 1st + 2nd KIND ?

1) RESULT OF EXPERIMENT

e.g. Is spin of resonance = 2 ?

GET ANSWER WRONG

Where to set χ^2 cut ?

Large cut : "Never" reject anything

Small cut : Reject when correct

Depends on nature of hypothesis

e.g. Does our result agree with that of expt E...?

OR Is our data consistent with Special Relativity ?

2) EVENT SELECTOR

e.g. Does this event contain Z^0 ?

Error of 1st kind : Loss of effic

Error of 2nd kind : Bgd

Usually easier to allow for 1 than 2.

3) PATTERN RECOGNITION

TRACKING

Hypothesis Testing = Pattern Recognition

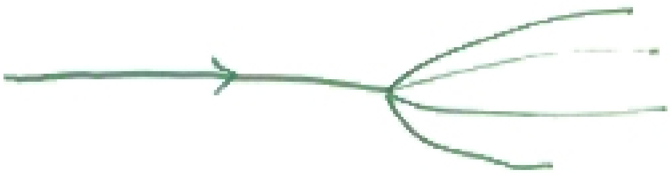
= Find hits that make track

Parameter Determination = Estimate track parameters

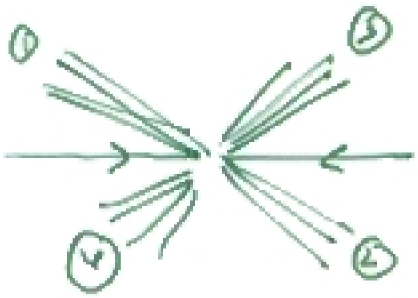
(+ error matrix)

KINEMATIC FITTING

Test whether observed event consistent with specified reaction

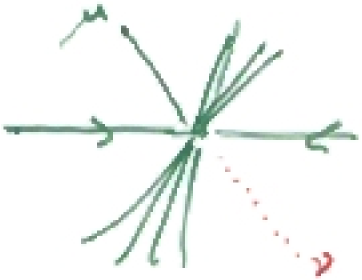
e.g. 

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



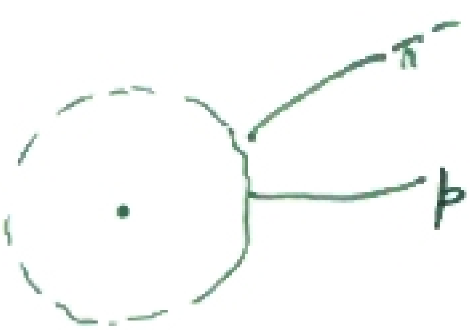
$e^+e^- \rightarrow W^+W^-$, $W^+ \rightarrow j_1 j_2$, $W^- \rightarrow j_3 j_4$

M_W, jet pairings



$e^+e^- \rightarrow W^+W^-$, $W^+ \rightarrow j_1 j_2$, $W^- \rightarrow \mu\nu$

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



p + π^- intersect

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

WHY DO IT ?

1) CHECK WHETHER EVENT CONSISTENT WITH HYPOTHESIS [HYPOTHESIS TESTING]

2) CAN CALCULATE MISSING VARIABLES [PARAM DETN.]

3) GOOD TO HAVE TRACKS CONSERVING E-P [P.D]

4) IMPROVES ERRORS [P.D] !

WHY DO IT ?

1) CHECK WHETHER EVENT CONSISTENT WITH HYPOTHESIS [HYPOTHESIS TESTING]

Use S_{min} & No of ~~constraints~~ degrees of freedom

2) CAN CALCULATE MISSING VARIABLES [PARAM DETN.]

e.g |P| for straight / short track / incoming ν

3 momentum of n, ν, ...

3) GOOD TO HAVE TRACKS CONSERVING E-P [P.D]

e.g. Identical values for resonance mass from prodn or from decay

4) IMPROVES ERRORS [P.D] !

Example of

"Adding theoretical Input can improve error"

$p\bar{p} \to p\bar{p}\,\pi^+\pi^-$ ✡

Measured variables

4 momenta of each track

(ie. 3 momenta + assumed/measured track identity)

Then test hypothesis:

Observed event = example of reaction ✡

Tested by:

Observed tracks should conserve E-p

Can tracks be "wiggled a bit" in order to do so?

$$\text{i.e. } S_{min} = \sum_{\substack{4 \text{ tracks} \\ \times 4 \text{ E-p}}} \left(\frac{v_i^{fitted} - v_i^{meas}}{\sigma_i} \right)^2$$

← If uncorr. Otherwise use Inv. Err. Matrix

where v_i^{fitted} conserve 4-momenta

i.e. Minimisation subject to constraint

(involves Lagrange multipliers)

TOY EXAMPLE OF FIT

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

4 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 $\Rightarrow$ REDUCED ERRORS

$$\lambda = \frac{\theta_1^m + \theta_2^m - \pi/2}{\sigma^2}$$

PARADOX

Histogram with 100 bins

Fit with one parameter

S_{min} : χ^2 with NDF = 99 $\quad (\overline{\chi^2} = 99 \pm 14)$

For our data, $S_{min}(p_0) = 85$

Is p_1 acceptable if $S(p_1) = 110$?

1) YES

Very acceptable χ^2 probability

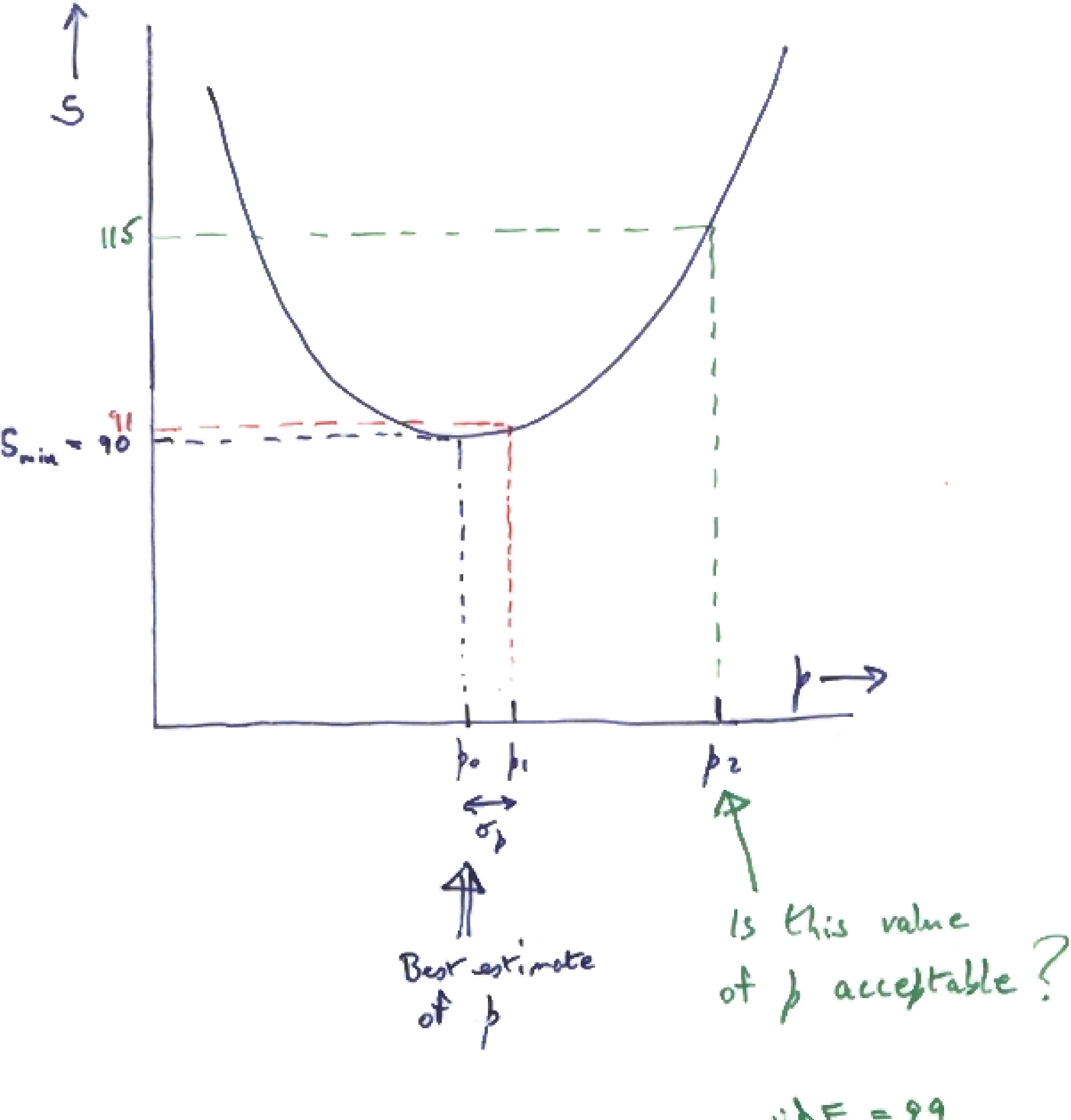
2) NO

σ_p from $\quad S(p_0 \pm \sigma_p) = S_{min} + 1$

$$= 86$$

But $S(p_1) - S_{min}(p_0) = 25$

$\therefore$ 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?
NDF = 99

2H 3

SELECTING BETWEEN TWO HYPOTHESES

LOUIS LYONS

OUNP-99-12

MATHEMATICAL FORMULATION

$$S(x) = \sum \frac{(x_i - x)^2}{\sigma^2} \equiv \sum \frac{(x_i - \bar{x})^2}{\sigma^2} + N\frac{(\bar{x} - x)^2}{\sigma^2}$$

SCATTER OF POINTS WRT THEIR MEAN.

INDEP OF x

THIS IS TERM WHICH HAS EXPECTED VALUE $(N-1) \pm \sqrt{2(N-1)}$

χ^2_{N-1}

HOW WELL x AGREES WITH $\bar{x}$

VARIES WITH x

BEST VALUE IS $x = \bar{x}$

INCREASES BY 1 FOR $x = \bar{x} \pm \frac{\sigma}{\sqrt{N}}$

χ^2_1

CONCLUSION FOR THIS CASE

COMPARING H1 : $p = p_1$

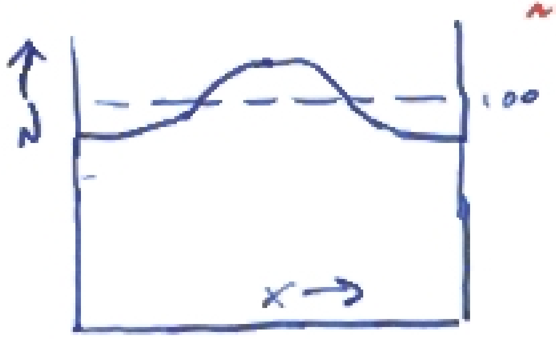
& H2 : $p = p_2$

DECISION DEPENDS ON $\Delta\chi^2$

ANOTHER EXAMPLE

$x = -1 \rightarrow +1$

100 bins
~100 events/bin
No FREE PARAMS



H1 : $1 + ax$
$a = 0.05$

H2 : $1 + b\cos(\pi x)$
$b = 0.05$

Generate events according to H1 (+ stat fluctn)

Try fitting according to H1 or to H2

χ^2_1 χ^2_2

Look at dist of χ^2_1 As expected for NDF=100

χ^2_2 Bit bigger. Many* "satisfactory"

$\chi^2_2 - \chi^2_1$ Decision based on $\Delta\chi^2$ has much better power

Repeat for events generated according to H2

Look at dist of χ^2_1

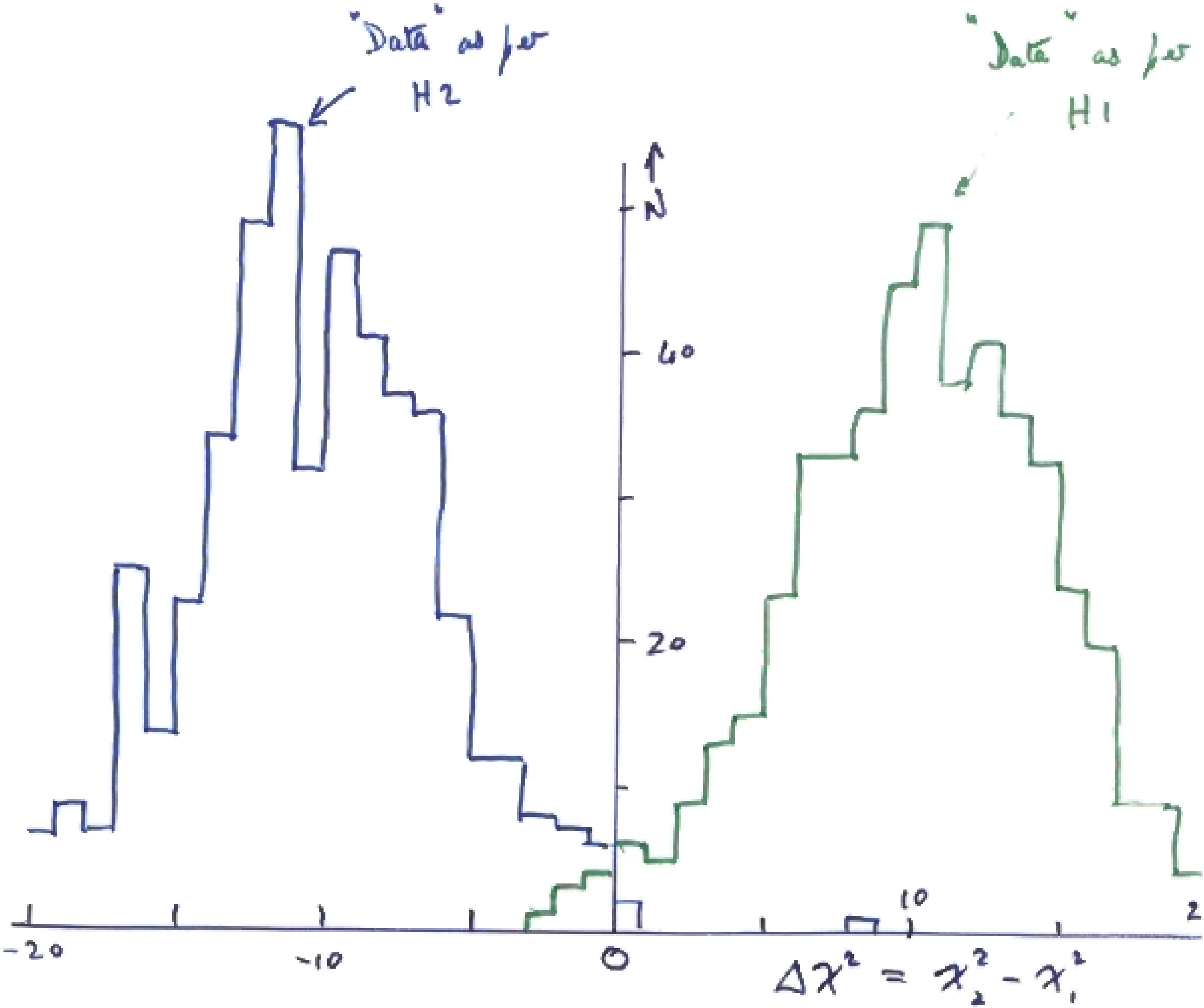
χ^2_2

$\chi^2_2 - \chi^2_1$

* 69% have $\chi^2_2 < 130$

DISTINGUISHING 2 HYPOTHESES ON BASIS OF $\Delta\chi^2$

(500 SIMULATIONS)



$H2 = 1 + 0.05 \cos(\pi x)$

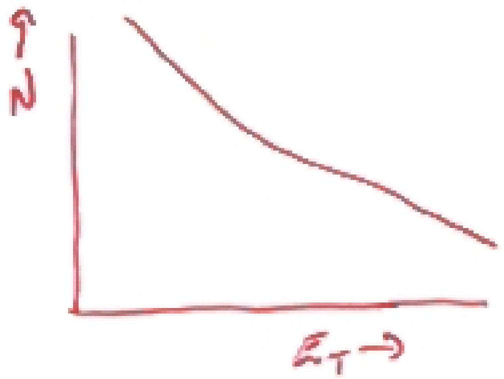
$H1 = 1 + 0.05 x$

BAYESIAN

2N 7

POSSIBLE APPLICATIONS

1) JET E_T DISTRIBUTION AT COLLIDER



FIT DISTRIBUTION BY ALTERNATIVE HYPOTHESES
(DIFFERENT STRUCTURE FNS.)

LOOK AT χ^2_1 FOR STR FN 1

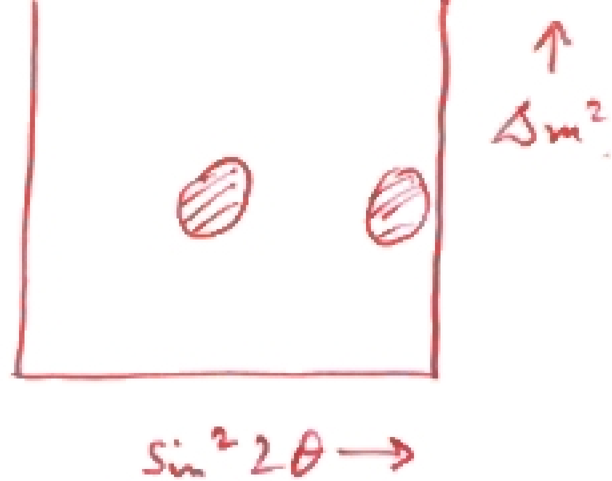
χ^2_2 FOR STR FN 2

DECIDE BETWEEN STR. FNS. ON BASIS OF

$\Delta\chi^2$

EVEN IF LARGER χ^2 GIVES O.K. PROB

2) ν OSCILLATIONS

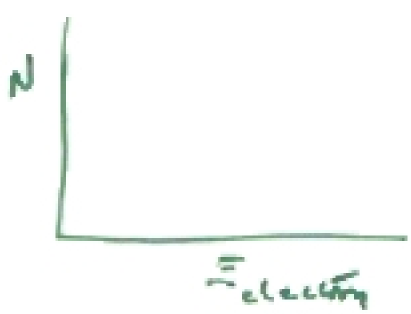


LARGE ANGLE } SOLNS FOR SOLAR NEUTRINOS
SMALL ANGLE }

(FROM OVERALL RATES FROM DIFFERENT DETECTORS)

LOOK AT DISTRIBUTION OF EVENTS IN SOME VARIABLE

e.g.



or Night/Day time

or Earth → Sun distance

(Different solns give slightly different predns)

Assume distribution has χ_1^2 for soln 1
χ_2^2 - - - - 2

Choose between solns on basis of $\Delta\chi^2$,

even if larger χ^2 has OK probability ..