χ^2 and Goodness of Fit

Louis Lyons

Oxford

CERN, October 2006

Lecture 3

Least squares best fit

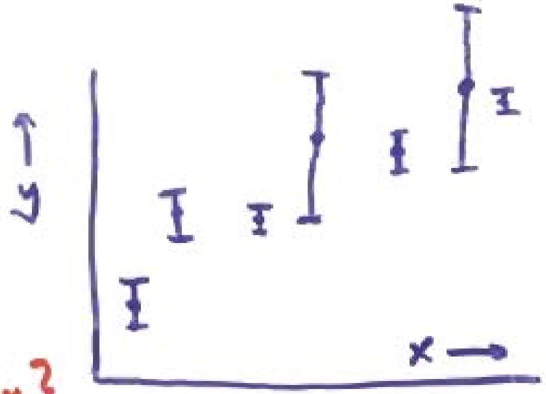
- Resume of straight line
- Correlated errors
- Errors in x and in y

Goodness of fit with χ^2

- Errors of first and second kind
- Kinematic fitting
  - Toy example

THE paradox

LEAST SQUARES STRAIGHT LINE FITTING



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

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.

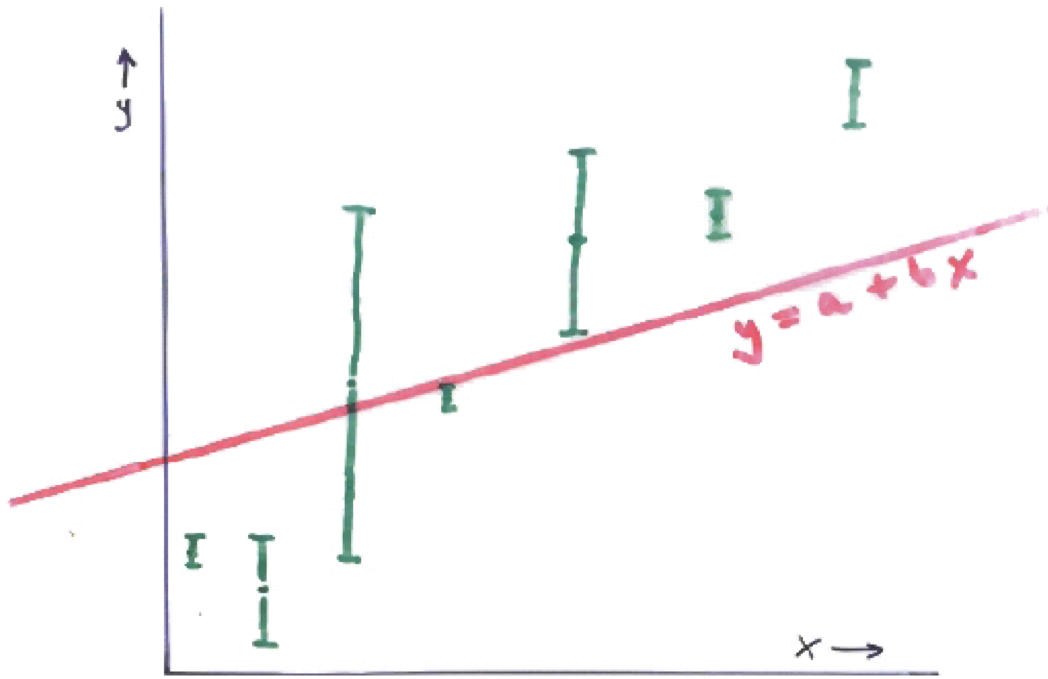
$y^{th} \sim y^{obs} \implies S$ small

Minimise $S \Rightarrow$ best line

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

*

Th	Obs	σ_{th}	σ_{obs}	Contr 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 $a_1 \pm \sigma_1$, $a_2 \pm \sigma_2$, $\vdots$, $a_i \pm \sigma_i$ } Best value $\hat{a} \pm \sigma$

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

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

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

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

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

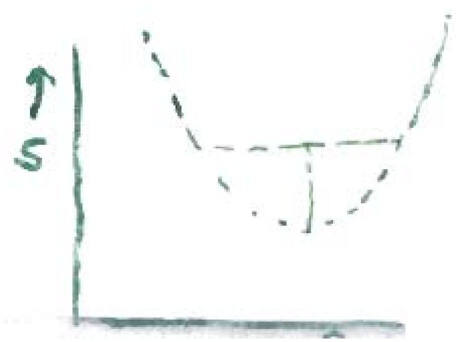
IN PARABOLIC APP EQUIV TO $S \to S_{min} + 1$

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

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

Many params

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



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 ($\underline{\underline{a}}$ & $\underline{\underline{b}}$)

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

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

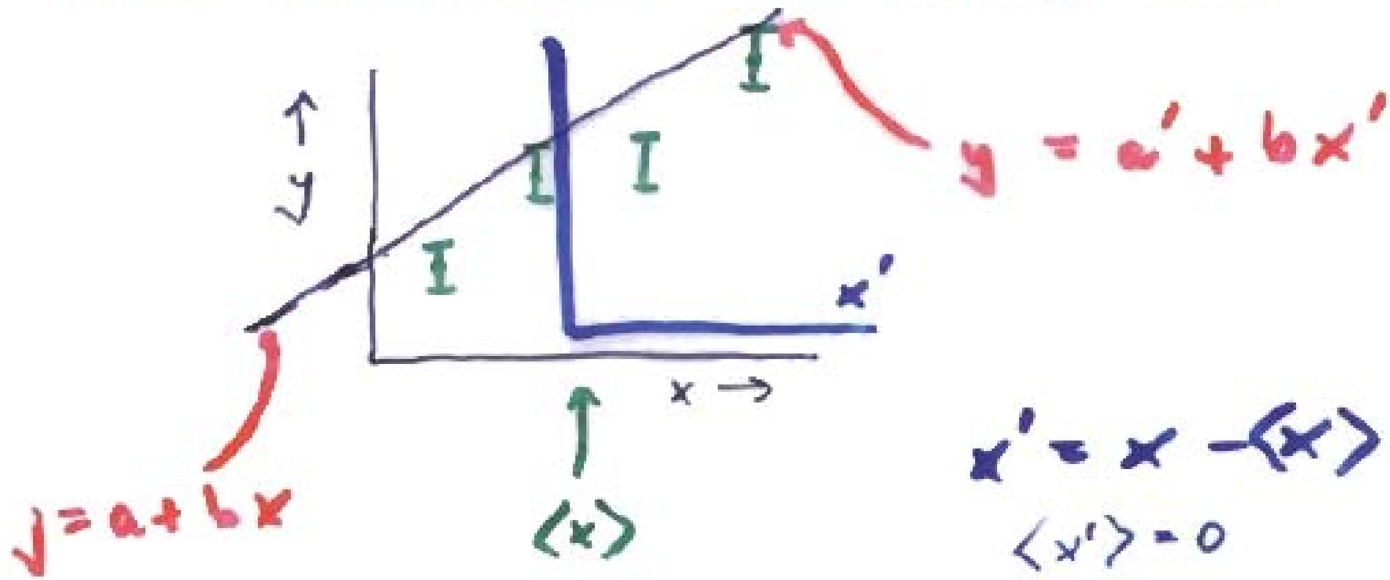
& $\langle f \rangle = [f]/[1]$

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

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

Error on intercept and gradient

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

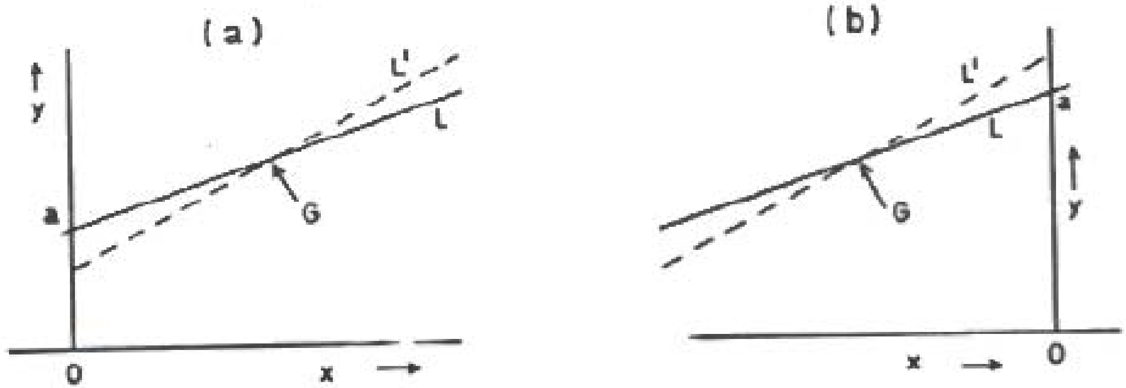


Better to use x' because error on a' & b are UNCORRELATED

[cf. Errors on a & b CORRELATED]

That is why track parameters specified at track 'centre'

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

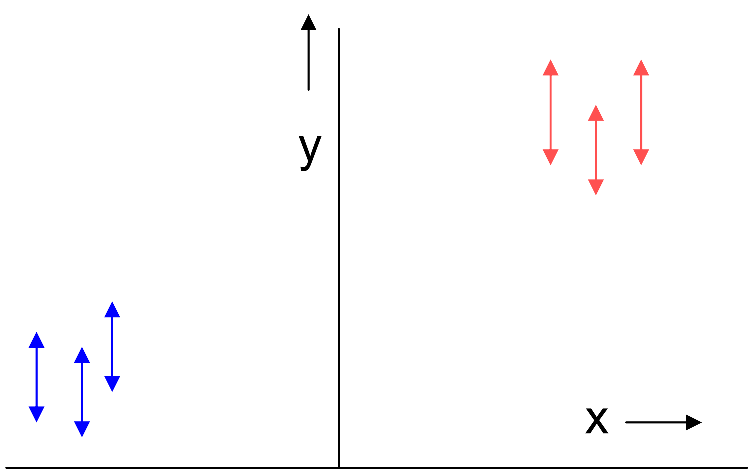


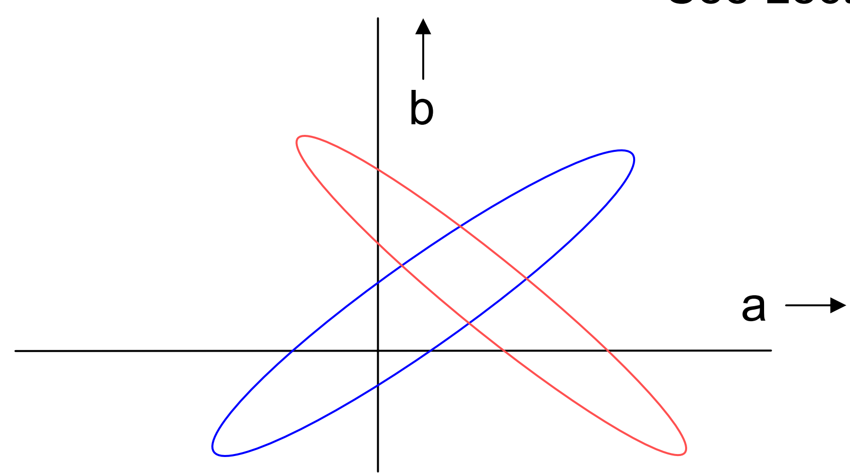
$\langle x \rangle$ pos

Fig. 2.4

$\langle x \rangle$ neg

See Lecture 1





If no errors specified on y_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 * ★

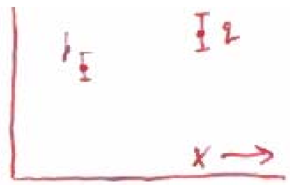
8) IF PROBABILITY TOO SMALL, IGNORE RESULTS

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

* COMPUTER PROGRAMME

★ Asymptotically

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

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

Inv. err matrix element →

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

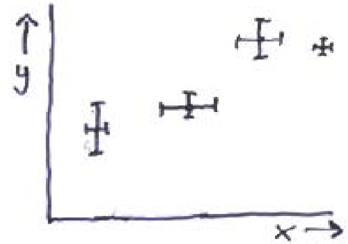
where $H^{-1} = \begin{pmatrix} \sigma_r^2 & cov \\ cov & \sigma_s^2 \end{pmatrix}$ ← Error matrix

Reduces to standard formula in absence of corrlns

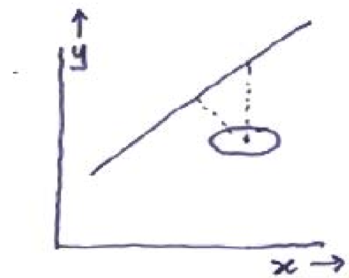
In general : $S = \sum_{ij} \tilde{\Delta}_i H_{ij} \Delta_j$

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

STRAIGHT LINE: Errors on x and on y



For simplicity,
assume x, y errors uncorrelated



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

Now replace by

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

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 $S = \sum \dfrac{(y_i - a - b\,x_i)^2}{\sigma_{y_i}^2 + b^2\sigma_{x_i}^2}$

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

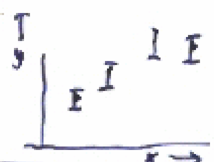
4) Can calculate χ^2 "on line" (i.e. single pass through data)

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

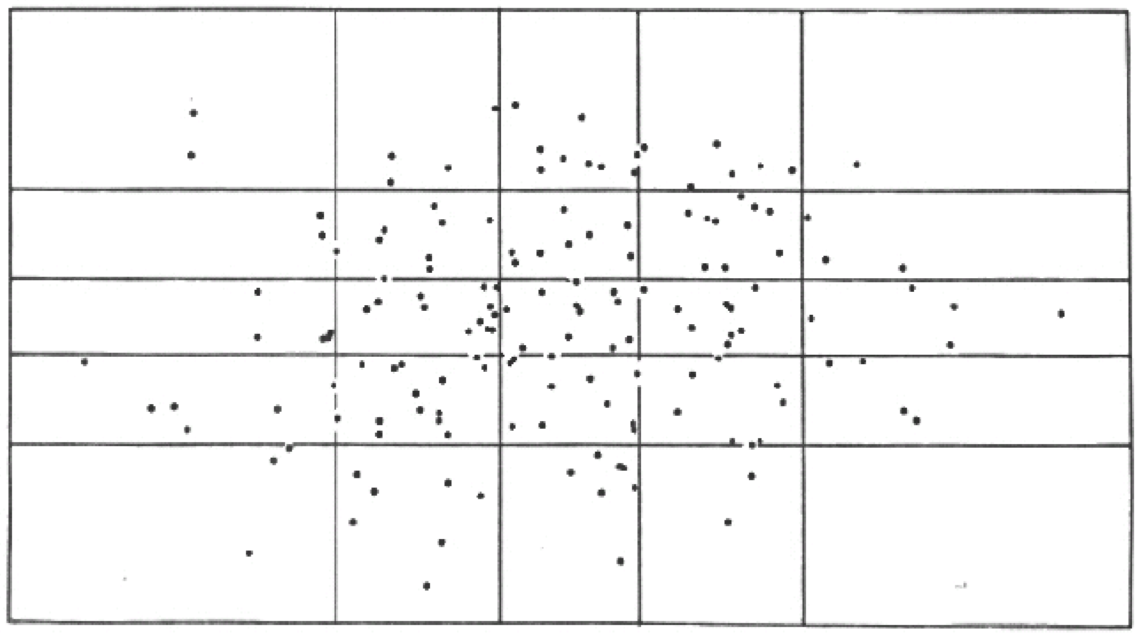
5) For theory linear in parameters,

soln can be found analytically

6) Hypothesis Testing ✡ ✡ ✡



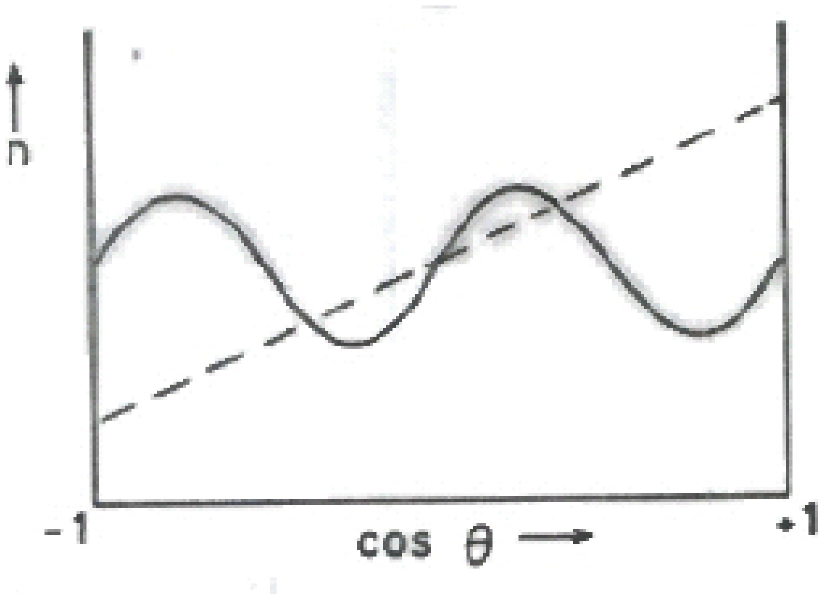
	Individual events (e.g. in $\cos\theta$)	$y_i \pm \sigma_i$ v x_i (e.g. stats)
1) Binning first	✓	×
4) χ^2 on line	First histogram	✓



	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	$\left(-\frac{\partial^2 \ell}{\partial p_i \partial p_j}\right)^{-\frac{1}{2}}$	$\left(\frac{1}{2}\frac{\partial^2 S}{\partial p_i \partial p_j}\right)^{-1}$
Main +	EASY	BEST FEW EVENTS	HYP. TEST.

‘Goodness of Fit’ by parameter testing?

$1+(b/a)\cos^2\theta$ Is $b/a = 0$?



‘Distribution testing’ is better

Goodness of Fit

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, $\chi^2 \geq S_{min}$

Works asymptotically , otherwise MC

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

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

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

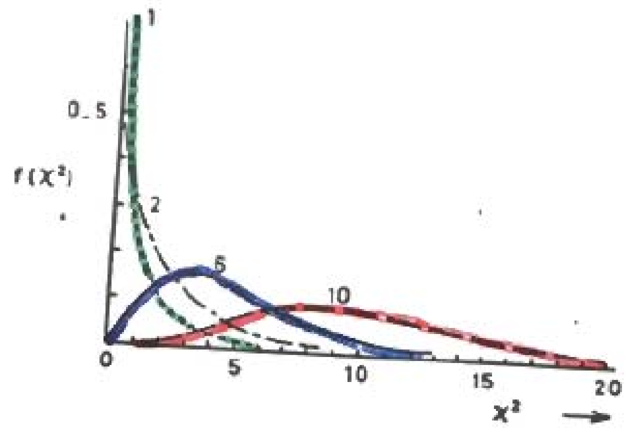


Fig. 2.6

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

is $\underline{\underline{\text{LARGE}}}$

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

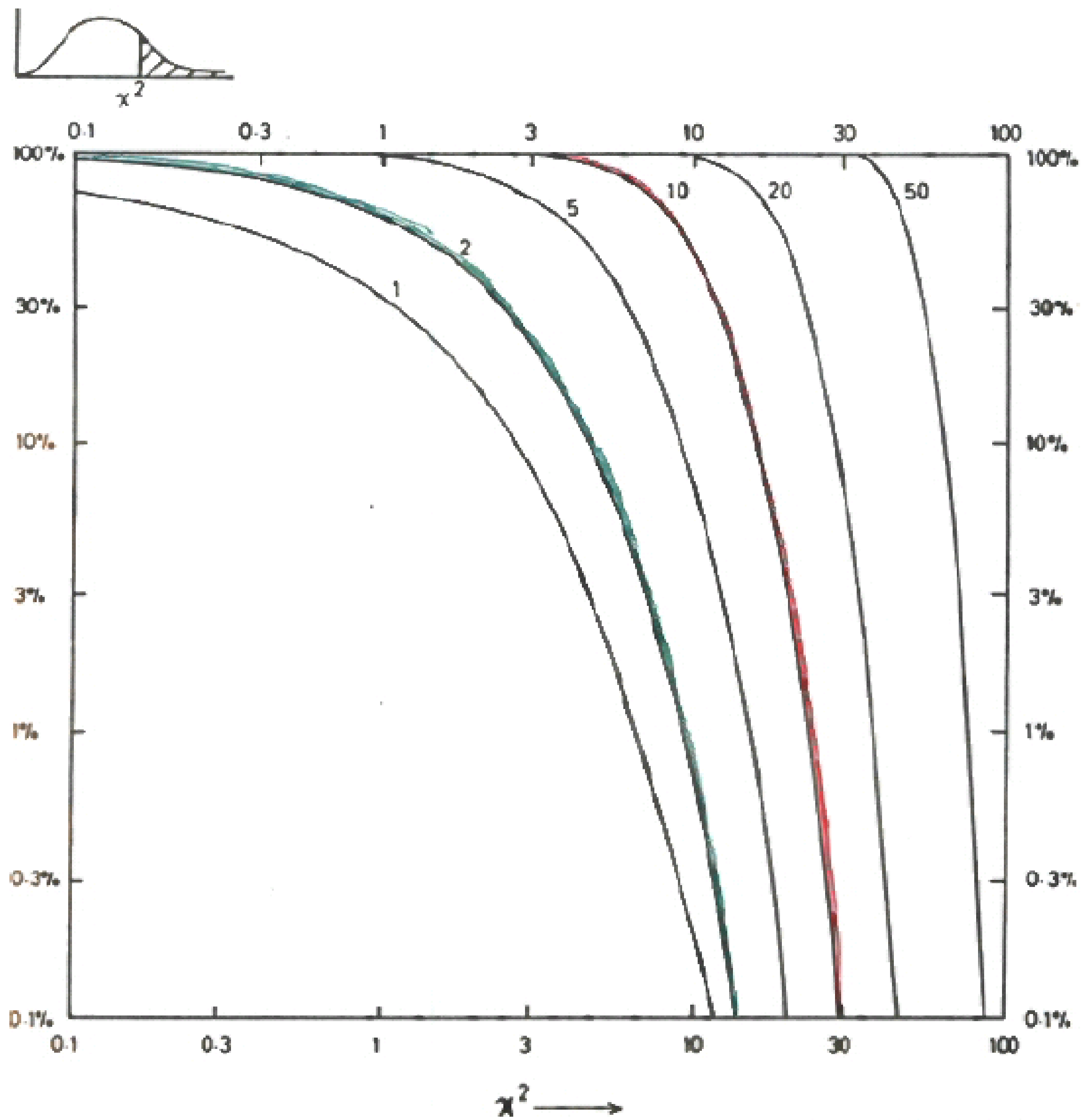


Fig. 2.7

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

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

b) Neutrino oscillations: almost degenerate parameters

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

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

Goodness of Fit

χ^2 : Very general

Needs binning

Not sensitive to sign of devn.

Run test

Kolmogorov - Smirnov

etc

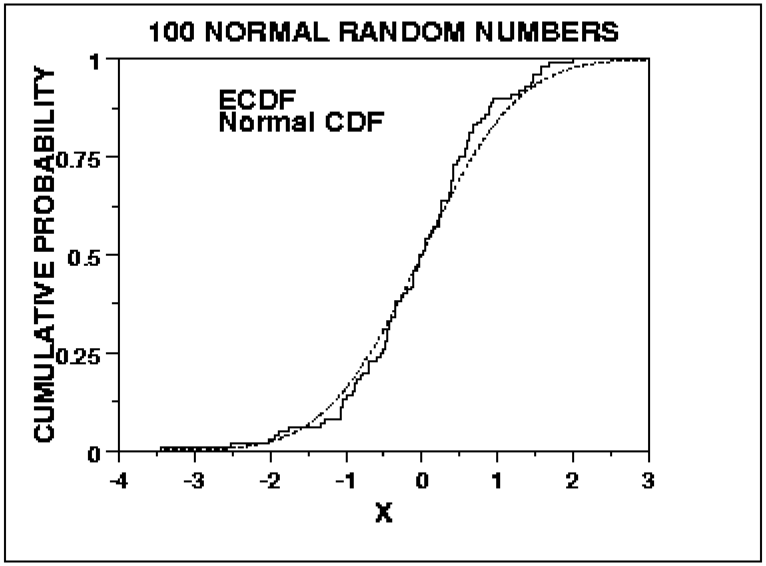
See: Aslan + Zech, Durham IPPP

Statistics Conf (2002)

Maria Grazia Pia's group in Genoa

Goodness of Fit: Kolmogorov-Smirnov

Compares data and model cumulative plots
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

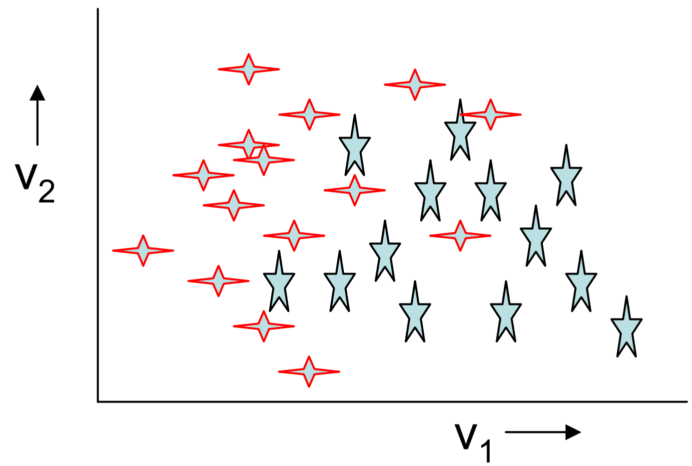
Assign +ve charge to data ; -ve charge to M.C.

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 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 μ K p ...

ii) Relative frequencies

e.g. 10^{-4} 10^{-4} 10% 10%

Aim for maximum effic ← Small error 1st kind

maximum purity ← small error 2nd kind

As χ^2 cut increases, effic ↑ purity ↓

Choose compromise

HOW SERIOUS ARE ERRORS OF
1^{st} + 2^{nd} 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

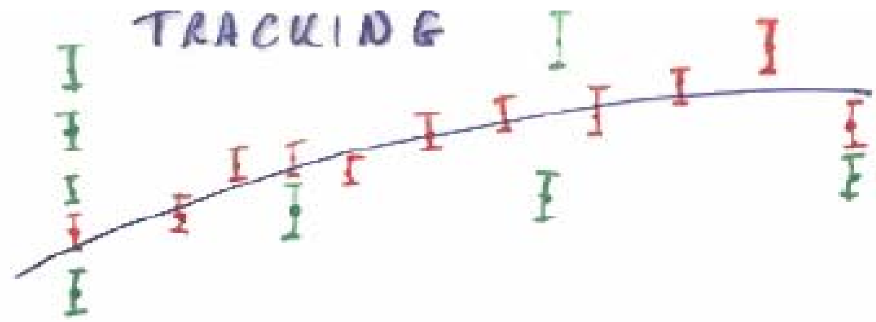
e.g. Does this event contain Z^0 ?

Error of 1^{st} kind : Loss of effic

Error of 2^{nd} kind : Bgd

Usually easier to allow for 1 than 2.

3) PATTERN RECOGNITION



Goodness of Fit: = Pattern Recognition

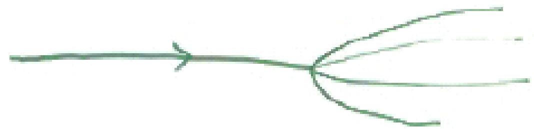
= Find hits that belong to track

Parameter Determination = Estimate track parameters

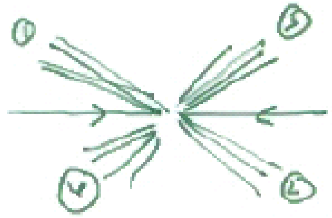
(and error matrix)

KINEMATIC FITTING

Test whether observed event consistent with specified reaction

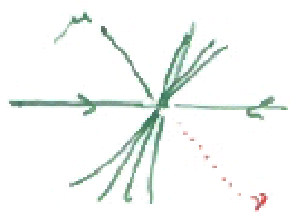
e.g. 

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



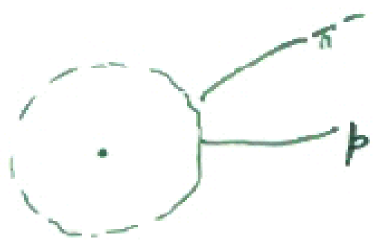
$e^+e^- \rightarrow W^+W^-$, $W^+ \hookrightarrow j_1 j_2$, $W^- \rightarrow j_3 j_4$

M_W, jet pairings



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

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



$p + \pi^-$ intersect

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

Kinematic Fitting: 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]

Kinematic Fitting: Why do it?

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

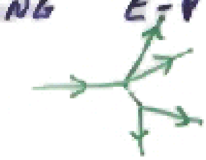
Use S_{min} & No of ~~constraints~~ degrees of freedom

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"

Measured variables $\qquad p\bar{p} \to p\bar{p}\,\pi^+\pi^-$ ✡

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 ?

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

← 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 Kinematic Fit

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

+ constraints:

1) Coplanar
2) p_1 at θ_1
3) p_2 at θ_2
4) θ_1 at θ_2 $\Leftarrow$ 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

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, $S_{min}(p_0) = 90$

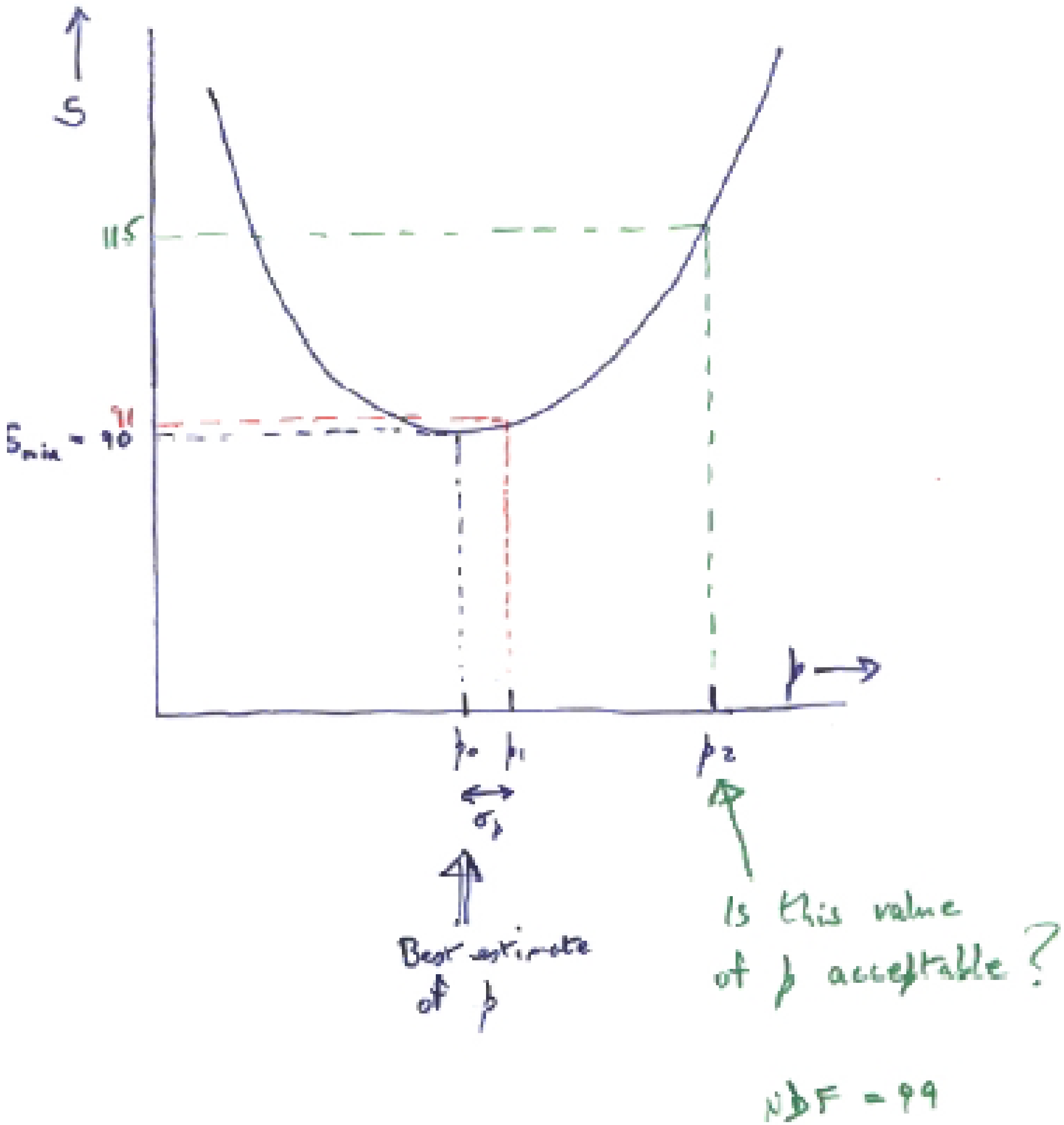
Is p_1 acceptable if $S(p_1) = 115$?

1) YES. Very acceptable χ^2 probability

2) NO. σ_p from $S(p_0 + \sigma_p) = S_{min} + 1 = 91$

But $S(p_1) - S(p_0) = 25$

So p_1 is 5σ away from best value


S
115
91
Smin = 90
p
p0
p1
p2
σp
Best estimate
of p
Is this value
of p acceptable?
NDF = 99

Next time:

Bayes and Frequentism: the return of an old controversy

The ideologies, with examples
Upper limits
Feldman and Cousins
Summary