Practical Statistics for Physicists

Louis Lyons

Imperial College and Oxford

CDF experiment at FNAL

CMS expt at LHC

l.lyons@physics.ox.ac.uk

Padova

May 2010

Topics

1) Introduction; +

   Learning to love the Error Matrix

2) Do's and Dont's with $\mathcal{L}$ikelihoods

3) Discovery and p-values

Time for discussion

Some of the questions to be addressed

What is coverage, and do we really need it?

Should we insist on at least a 5σ effect to claim discovery?

How should p-values be combined?

If two different models both have respectable χ^2 probabilities, can we reject one in favour of other?

Are there different possibilities for quoting the sensitivity of a search?

How do upper limits change as the number of observed events becomes smaller than the predicted background?

Combine 1 ± 10 and 3 ± 10 to obtain a result of 6 ± 1?

What is the Punzi effect and how can it be understood?

Books

Statistics for Nuclear and Particle Physicists

Cambridge University Press, 1986

Available from CUP

Errata in these lectures

Other Books

J. OREAR "NOTES ON STATISTICS FOR PHYSICISTS
UCRL-8417 (1958)

D J HUDSON 'Lectures on elementary statistics + prob.'
+ 'Max like + least squares theory'
CERN reports 63-29 + 64-1

S. BRANDT STATISTICAL & COMPUTATIONAL METHODS IN
DATA ANALYSIS (North Holland 1973)

WT EADIE et al STATISTICAL METHODS IN
EXPTL PHYSICS (North Holland 1971)

SL MEYER DATA ANALYSIS FOR SCIENTISTS + ENGINEERS (Wiley 1975)

A FRODESON et al PROBABILITY + STATISTICS IN
PARTICLE PHYSICS (Bergen 1979)

R. BARLOW "STATISTICS (Wiley, 1993)

G COWAN, STATISTICAL DATA ANALYSIS (Oxford 1998)

B. ROE PROBABILITY & STATISTICS IN EXPTL PHYSICS
(Springer-Verlag 1992)

Particle Data Book

CDF Statistics Committee

BaBar Statistics Working Group

Frederick James

Statistical Methods in Experimental Physics

2nd Edition

Statistical Methods in Experimental Physics
2nd Edition

The first edition of this classic book has become the authoritative reference for physicists desiring to master the finer points of statistical data analysis. This second edition contains all the important material of the first, much of it unavailable from any other sources. In addition, many chapters have been updated with considerable new material, especially in areas concerning the theory and practice of confidence intervals, including the important Feldman–Cousins method. Both frequentist and Bayesian methodologies are presented, with a strong emphasis on techniques useful to physicists and other scientists in the interpretation of experimental data and comparison with scientific theories. This is a valuable textbook for advanced graduate students in the physical sciences as well as a reference for active researchers.

World Scientific
25 Years of Publishing
1981 – 2006

6008 hc
ISBN 981-256-795-X
9 789812 567956
www.worldscientific.com

2nd Edition
Statistical Methods in Experimental Physics
James

Introductory remarks

Probability and Statistics

Random and systematic errors

Conditional probability

Variance

Combining errors

Combining experiments

Binomial, Poisson and Gaussian distributions

PROBABILITY STATISTICS

Example : Dice

Given $P(5) = 1/6$, what is P(20 5's out of 100 trials)?

Given 20 5's out of 100, what is $P(5)$? And its error?

If unbiassed, what is P(n evens out of 100 trials)?

Observe 65 evens in 100 trials

Is it unbiassed?

Or is $P(\text{even}) = \frac{2}{3}$?

PROBABILITY

STATISTICS

Example : Dice

Probability

Given $P(5) = 1/6$, what is P(20 5's out of 100 trials)?

If unbiassed, what is P(n evens out of 100 trials)?

THEORY→ DATA

Statistics

Given 20 5's out of 100, what is $P(5)$? And its error?

Parameter Determination

Observe 65 evens in 100 trials

Is it unbiassed?

Goodness of fit

Or is $P(\text{even}) = \frac{2}{3}$?

Hypothesis testing

DATA→ THEORY

N.B. Parameter values not sensible if goodness of fit is poor/bad

Why do we need errors?

Affects conclusion about our result e.g.

Result / theory = 0.970

If 0.970 ± 0.050, data compatible with theory

If 0.970 ± 0.005, data incompatible with theory

If 0.970 ± 0.7, need better experiment

Historical experiment at Harwell testing General Relativity

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.

CONDITIONAL PROBABILITY

$$\text{Prob}[A+B] = \frac{N(A+B)}{N_{tot}} = \frac{N(A+B)}{N(B)} \cdot \frac{N(B)}{N_{tot}}$$

$$= P(A|B) \times P(B)$$

If A + B are independent, $P(A|B) = P(A)$

$$\Rightarrow P(A+B) = P(A) \times P(B) \quad , \quad A + B \text{ indep}$$

e.g. $P[\text{Rainy} + \text{Sunday}] = P(\text{rainy}) \times \frac{1}{7}$ INDEP

$P[\text{Rainy} + \text{December}] \neq P(\text{rainy}) \times \frac{1}{12}$ ~~INDEP~~

$P[E_e \text{ large} + E_\nu \text{ large}] \neq P(E_e \text{ large}) \times P(E_\nu \text{ large})$ ~~INDEP~~

$P[\text{Beam part 1 interacts} + \text{Beam part 2 interacts}]$

$= [P(\text{beam particle interacts})]^2$ INDEP

$$\text{Prob}[A+B] = \text{Prob}[A|B] \times \text{Prob}[B]$$

$$= \text{Prob}[B|A] \times \text{Prob}[A]$$

Bayes' Theorem

ESTIMATE OF VARIANCE

$$s^2 = \frac{1}{N-1} \sum (x_i - \bar{x})^2$$

UNBIASSED ESTIMATE OF σ^2

$$= \frac{N}{N-1} \left(\overline{x^2} - \bar{x}^2 \right)$$

USEFUL "ON LINE"

BUT can have numerical problems

For Gaussian x_i

error on $s = \frac{\sigma}{\sqrt{2(N-1)}}$

e.g. $N = 3 \Rightarrow 50\%$ error

$N = 51$ for 10% error

Combining errors

$$z = x - y$$

$$\delta z = \delta x - \delta y \qquad [1]$$

Why $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$? [2]

Combining errors

$$z = x - y$$

$$\delta z = \delta x - \delta y \qquad [1]$$

Why $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$? [2]

1) [1] is for specific δx, δy

Could be so on average ?

N.B. Mneumonic, not proof

2) $\sigma_z^2 = \overline{\delta z^2} = \overline{\delta x^2} + \overline{\delta y^2} - 2\,\overline{\delta x\,\delta y}$

$= \sigma_x^2 + \sigma_y^2$ provided…………..

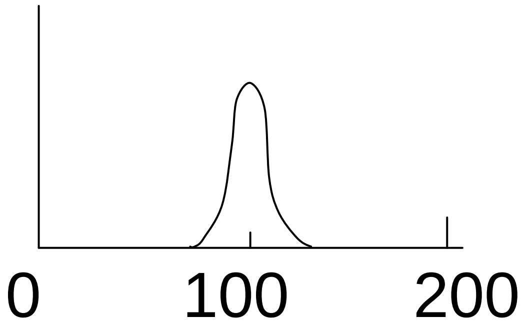
3) **Averaging is good for you**: N measurements $x_i \pm \sigma$

[1] $x_i \pm \sigma$ or [2] $x_i \pm \sigma/\sqrt{N}$?

4) **Tossing a coin**:

Score 0 for tails, 2 for heads (1 ± 1)

After 100 tosses, [1] 100 ± 100 or [2] 100 ± 10 ?



Prob(0 or 200) = $(1/2)^{99} \sim 10^{-30}$

Compare age of Universe $\sim 10^{18}$ seconds

Rules for different functions

1) Linear: $z = k_1x_1 + k_2x_2 + \ldots\ldots.$

$\sigma_z = k_1\,\sigma_1 \,\&\, k_2\,\sigma_2$

& means "combine in quadrature"

N.B. Fractional errors **NOT** relevant

e.g. $z = x - y$

z = your height

x = position of head wrt moon

y = position of feet wrt moon

x and y measured to 0.1%

z could be -30 miles

Rules for different functions

2) Products and quotients

$$z = x^{\alpha}\, y^{\beta}\ldots\ldots.$$

$$\sigma_z/z = \alpha\, \sigma_x/x \;\; \& \;\; \beta\, \sigma_y/y$$

Useful for x^2, xy, x/y,…….

3) Anything else:

$$z = z(x_1, x_2, \ldots..)$$

$$\sigma_z = \partial z/\partial x_1\ \sigma_1\ \&\ \partial z/\partial x_2\ \sigma_2\ \&\ \ldots\ldots.$$

OR numerically:

$z_0 = z(x_1, \quad x_2, \quad x_3 \ldots.)$

$z_1 = f(x_1+\sigma_1, x_2, \quad x_3 \ldots.)$

$z_2 = f(x_1, \quad x_2+ \sigma_2, \quad x_3 \ldots.)$

$\sigma_z = (z_1-z_0)\ \&\ (z_2-z_0)\ \&\ \ldots.$

N.B. All formulae approximate (except 1)) – assumes small errors

Introductory remarks

Probability and Statistics

Random and systematic errors

Conditional probability

Variance

Combining errors

Combining experiments

Binomial, Poisson and Gaussian distributions

COMBINING EXPERIMENTS

$x_i \pm \sigma_i$ (uncorrelated)

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

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

← Minimise S

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

← σ from $S_{min} + 1$

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

$$\left.\begin{array}{c} 100 \pm 10 \\ 1 \pm 1 \end{array}\right\} \Longrightarrow 2 \pm 1\ ?$$

or 50.5 ± 5 ?

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 $= 100 \pm 5$ K

Number of married women $= 80 \pm 30$ K

Total $= 180 \pm 30$ K

Wtd average $= 99 \pm 5$ K CONTRAST

Total $= 198 \pm 10$ K

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

Compare “kinematic fitting”

Binomial Distribution

Fixed N independent trials, each with same prob of success p

What is prob of s successes?

e.g. Throw dice 100 times. Success = '6'. What is prob of 0, 1,…. 49, 50, 51,… 99, 100 successes?

Effic of track reconstrn = 98%. For 500 tracks, prob that 490, 491,...... 499, 500 reconstructed.

Ang dist is $1 + 0.7 \cos\theta$? Prob of 52/70 events with $\cos\theta > 0$?

(More interesting is statistics question)

$$P_s = \frac{N!}{(N-s)!\, s!} p^s (1-p)^{N-s}$$, as is obvious

Expected number of successes = $\Sigma n P_n = Np$,

as is obvious

Variance of no. of successes = $Np(1-p)$

Variance ~ Np, for $p \sim 0$

~ $N(1-p)$ for $p \sim 1$

NOT Np in general. NOT $n \pm \sqrt{n}$

e.g. 100 trials, 99 successes, NOT 99 ± 10

Statistics: Estimate p and σ_p from s (and N)

$p = s/N$

$\sigma_p^2 = 1/N \; s/N \; (1 - s/N)$

If $s = 0$, $p = 0 \pm 0$?

If $s = 1$, $p = 1.0 \pm 0$?

Limiting cases:

- p = const, $N \rightarrow \infty$: Binomial $\rightarrow$ Gaussian

  $\mu = Np$, $\sigma^2 = Np(1-p)$

- $N \rightarrow \infty$, $p \rightarrow 0$, Np = const: Bin $\rightarrow$ Poisson

  $\mu = Np$, $\sigma^2 = Np$

{N.B. Gaussian continuous and extends to $-\infty$}

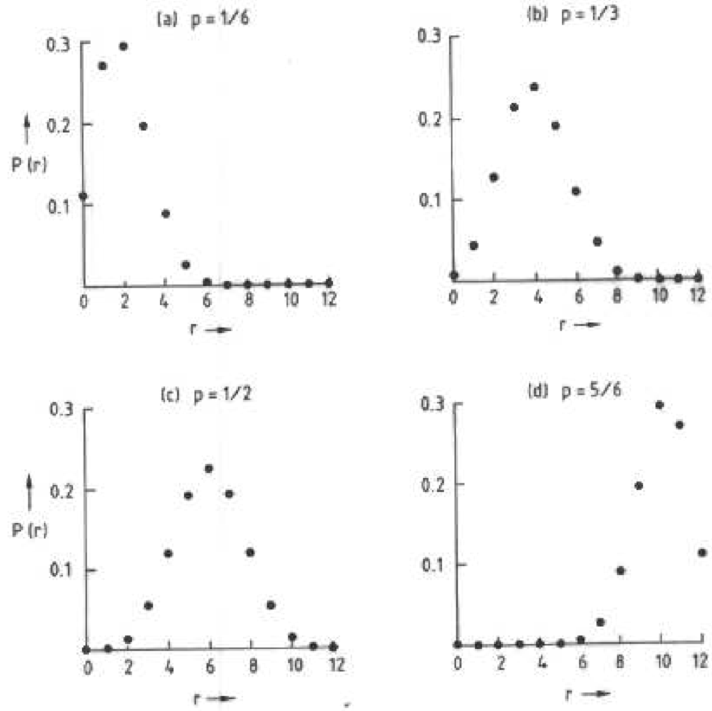


Fig. A3.1 The probabilities $P(r)$, according to the binomial distribution, for r successes out of 12 independent trials, when the probability p of success in an individual trial is as specified in the diagram. As the expected number of successes is $12p$, the peak of the distribution moves to the right as p increases. The RMS width of the distribution is $\sqrt{12p(1-p)}$ and hence is largest for $p = \frac{1}{2}$. Since the chance of success in the $p = \frac{1}{6}$ case is equal to that of failure for $p = \frac{5}{6}$, the diagrams (a) and (d) are mirror images of each other. Similarly the $p = \frac{1}{2}$ situation shown in (c) is symmetric about $r = 6$ successes.

Poisson Distribution

Prob of n independent events occurring in time t when rate is r (constant)

e.g. events in bin of histogram

NOT Radioactive decay for $t \sim \tau$

Limit of Binomial ($N \rightarrow \infty$, $p \rightarrow 0$, $Np \rightarrow \mu$)

$P_n = e^{-r\,t}\,(r\,t)^n / n! \quad = e^{-\mu}\,\mu^n / n! \quad (\mu = r\,t)$

$\langle n \rangle = r\,t = \mu$ (No surprise!)

$\sigma^2_n = \mu$ "$n \pm \sqrt{n}$"   BEWARE 0 ± 0 ?

$\mu \rightarrow \infty$: Poisson$\rightarrow$ Gaussian, with mean = μ, variance =μ

Important for χ^2

For your thought

Poisson $P_n = e^{-\mu} \mu^n/n!$

$P_0 = e^{-\mu}$ $P_1 = \mu\, e^{-\mu}$ $P_2 = \mu^2/2\ e^{-\mu}$

For small μ, $P_1 \sim \mu$, $P_2 \sim \mu^2/2$

If probability of 1 rare event $\sim \mu$,

why isn't probability of 2 events $\sim \mu^2$?

$$P_2 = \mu^2 e^{-\mu} \quad \text{or} \quad P_2 = \mu^2/2 \ e^{-\mu} \ ?$$

1) $P_n = e^{-\mu} \mu^n/n!$ sums to unity

2) n! comes from corresponding Binomial factor N!/{s!(N-s)!}

3) If first event occurs at t_1 with prob μ, average prob of second event in $t-t_1$ is $\mu/2$. (Identical events)

4) Cow kicks and horse kicks, each producing scream. Find prob of 2 screams in time interval t, by P_2 and P_2

2c, 0h	$(c^2e^{-c})\ (e^{-h})$	$(\tfrac{1}{2}\, c^2e^{-c})\ (e^{-h})$
1c,1h	$(ce^{-c})\ (he^{-h})$	$(ce^{-c})\ (he^{-h})$
0c,2h	$(e^{-c})\ (h^2e^{-h})$	$(e^{-c})\ (\tfrac{1}{2}\, h^2e^{-h})$
Sum	$(c^2 + hc + h^2)\ e^{-(c+h)}$	$\tfrac{1}{2}(c^2 + 2hc + h^2)\ e^{-(c+h)}$
2 screams	$(c+h)^2\ e^{-(c+h)}$	$\tfrac{1}{2}(c+h)^2\ e^{-(c+h)}$
	Wrong	OK

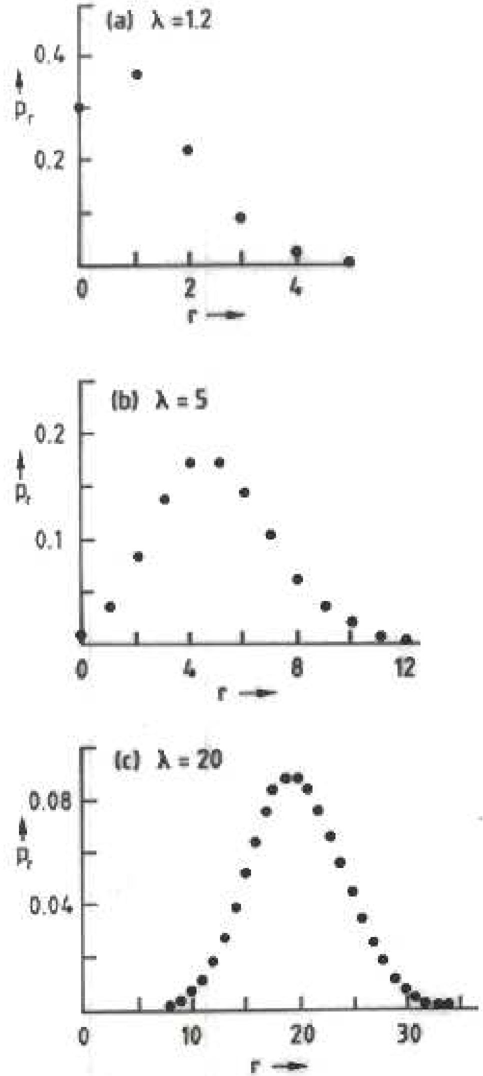


Fig. A4.1 Poisson distributions for different values of the parameter λ. (a) $\lambda = 1.2$; (b) $\lambda = 5.0$; (c) $\lambda = 20.0$. P_r is the probability of observing r events. (Note the different scales on the three figures.) For each value of λ, the mean of the distribution is at λ, and the RMS width is $\sqrt{\lambda}$. As λ increases above about 5, the distributions look more and more like Gaussians.

Relation between Poisson and Binomial

N people in lecture, m males and f females $(N = m + f)$

Assume these are representative of basic rates: ν people νp males $\nu(1-p)$ females

Probability of observing N people = $P_{Poisson} = e^{-\nu}\,\nu^{N}/N!$

Prob of given male/female division = $P_{Binom} = \frac{N!}{m!\,f!}\,p^{m}\,(1-p)^{f}$

Prob of N people, m male and f female = $P_{Poisson}\,P_{Binom}$

$$= \frac{e^{-\nu p}\,\nu^{m}\,p^{m}}{m!} \quad * \quad \frac{e^{-\nu(1-p)}\,\nu^{f}\,(1-p)^{f}}{f\,!}$$

= Poisson prob for males * Poisson prob for females

People	Male	Female
Patients	Cured	Remain ill
Decaying nuclei	Forwards	Backwards
Cosmic rays	Protons	Other particles

Gaussian or Normal

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

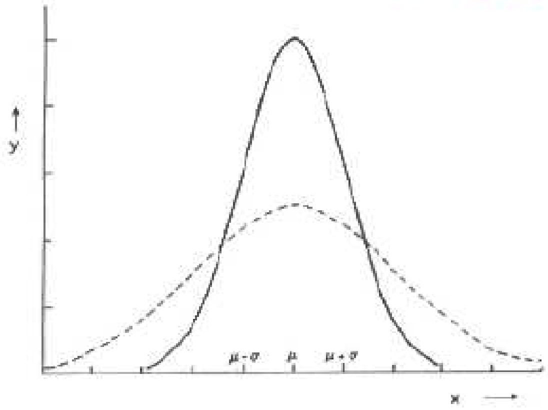


Fig. 1.5. The solid curve is the Gaussian distribution of eqn (1.14). The distribution peaks at the mean μ, and its width is characterised by the parameter σ. The dashed curve is another Gaussian distribution with the same values of μ, but with σ twice as large as the solid curve. Because the normalisation 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.

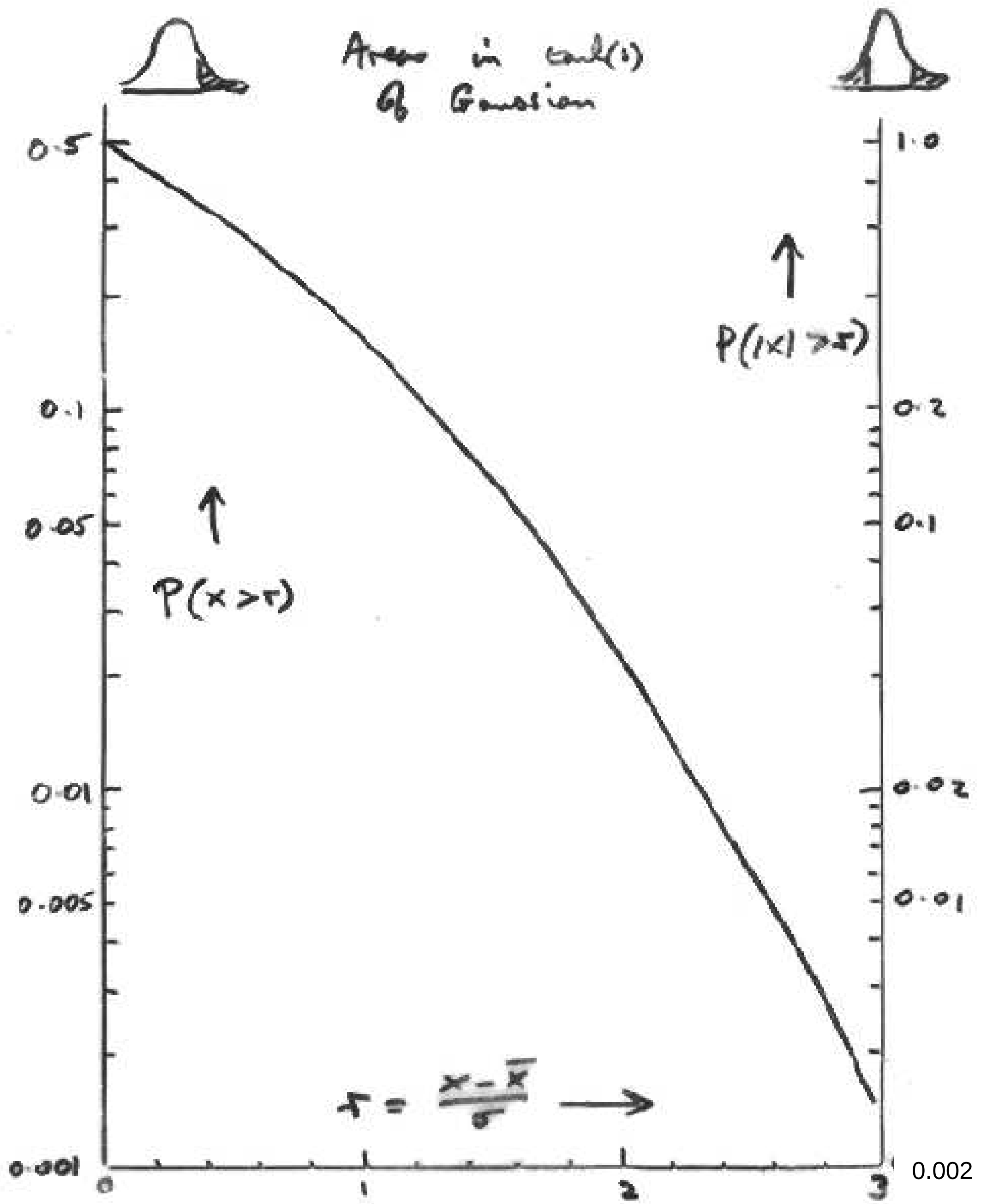
Significance of σ

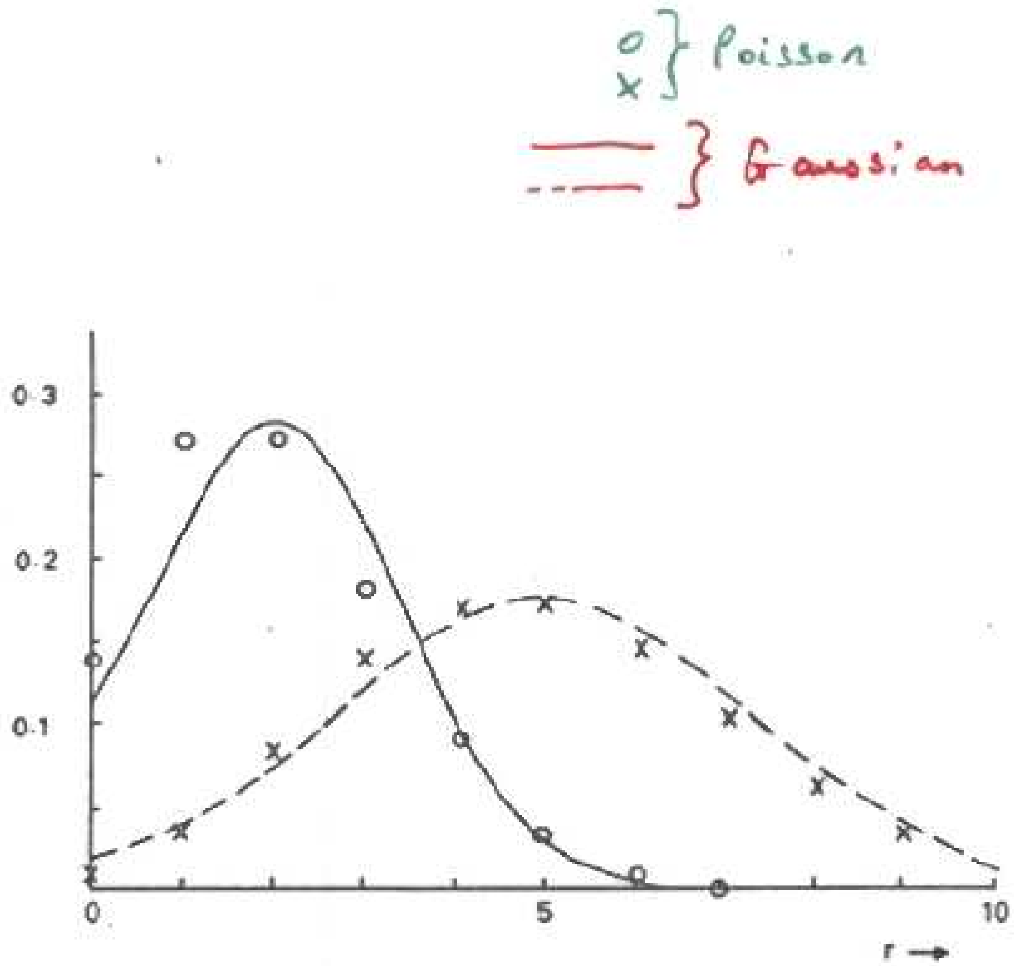
i) RMS of Gaussian $= \sigma$
(Hence factor of 2 in defn of Gaussian)

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

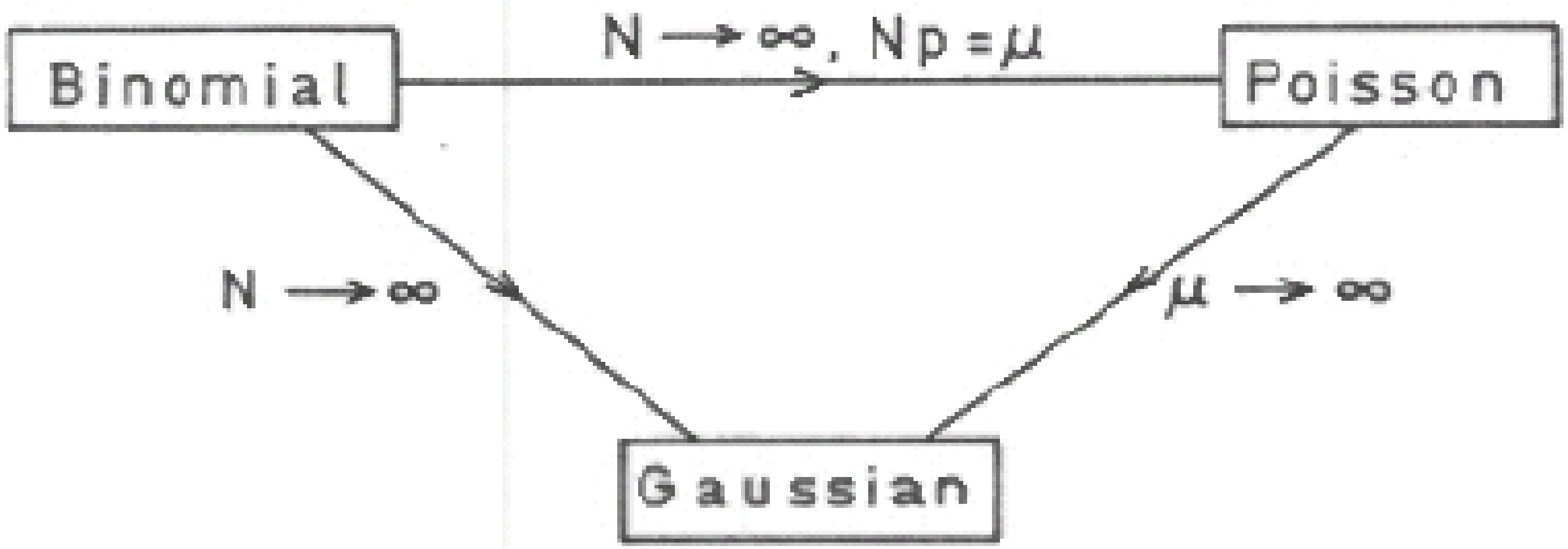
iii) Fractional area within $\mu \pm \sigma$ is 68%

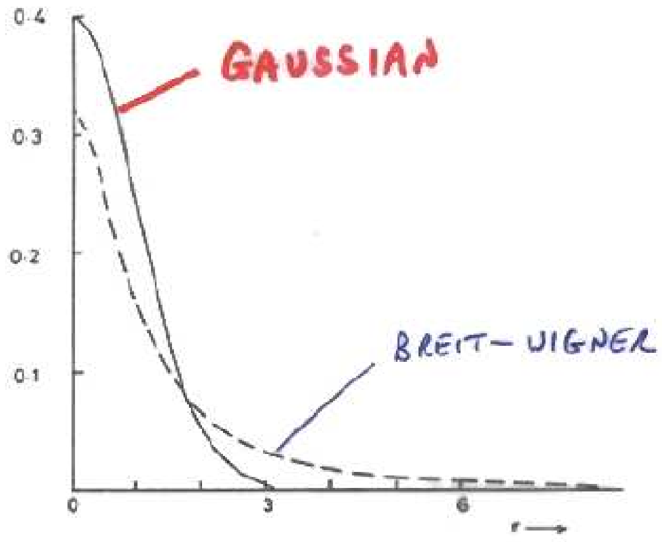
iv) Height at max $= 1/\sqrt{2\pi}\,\sigma$


Areas 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 > τ)
P(|x| > τ)
τ = (x − x̄)/σ
0
1
2
3



Relevant for Goodness of Fit


Binomial
$N \rightarrow \infty$, $Np = \mu$
Poisson
$N \rightarrow \infty$
$\mu \rightarrow \infty$
Gaussian



Gaussian = N (r, 0, 1)

Breit Wigner = $1/\{\pi * (r^2 + 1)\}$

STUDENT'S t

$$x = \frac{\text{meas} - \text{expected}}{\sigma}$$

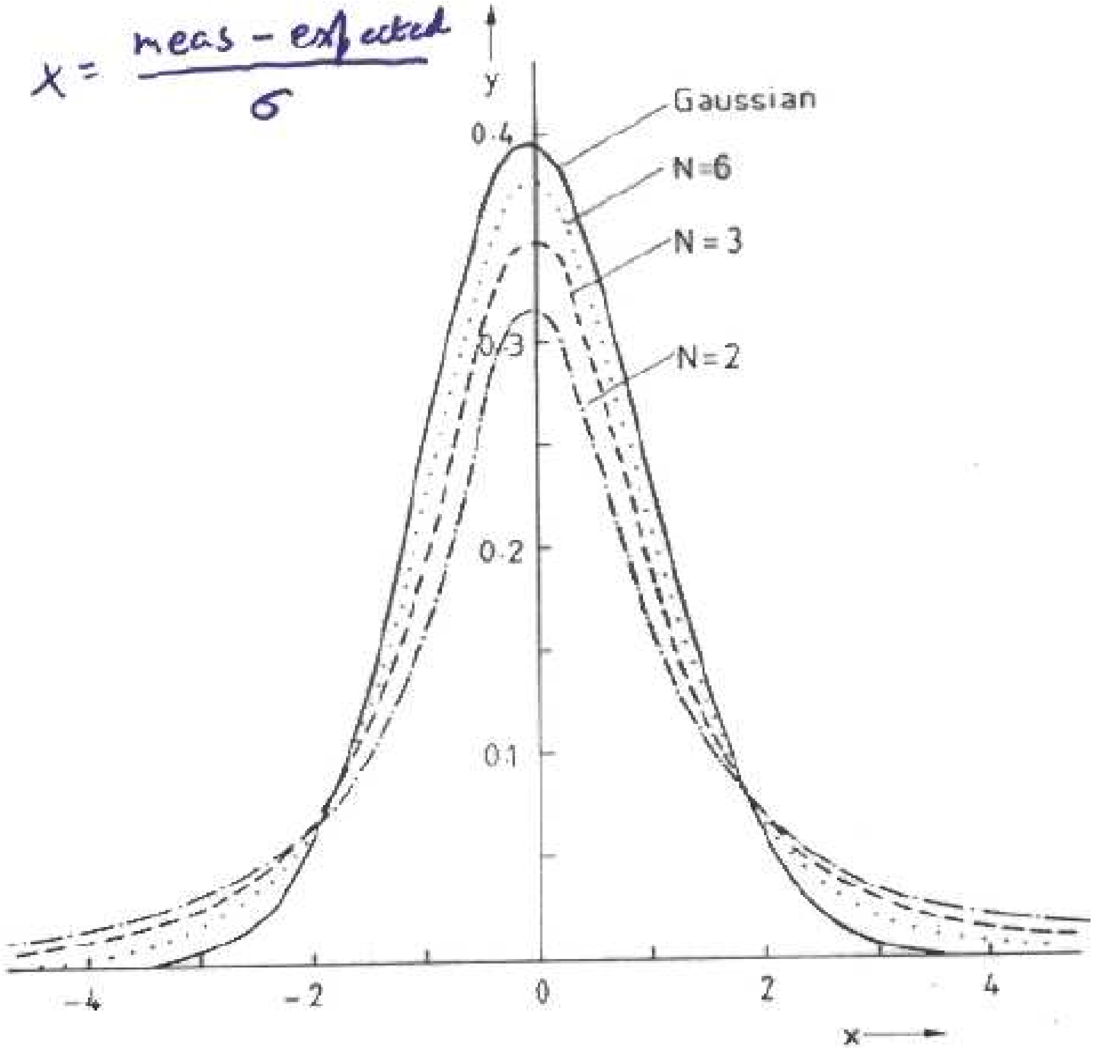


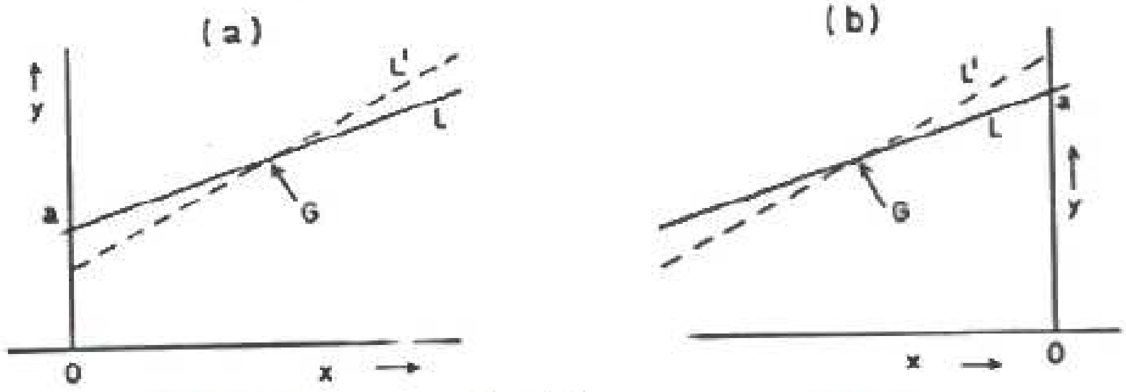
Fig. A5.1 Comparison of Student's t distributions for various values of the number of observations N, with the Gaussian distribution, which is the limit of the Student's distributions as N tends to infinity.

Learning to love the Error Matrix

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

  Combining correlated measurements
- Estimating the error matrix

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 more simple to introduce concept this way

<u>Gaussian in 2-variables</u>

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

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

$x + y$ uncorrelated $\Rightarrow$

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

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

$\Rightarrow$ ERROR MATRIX

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

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

Diagonal E_{ij} = variances

Off-diagonal E_{ij} = covariances

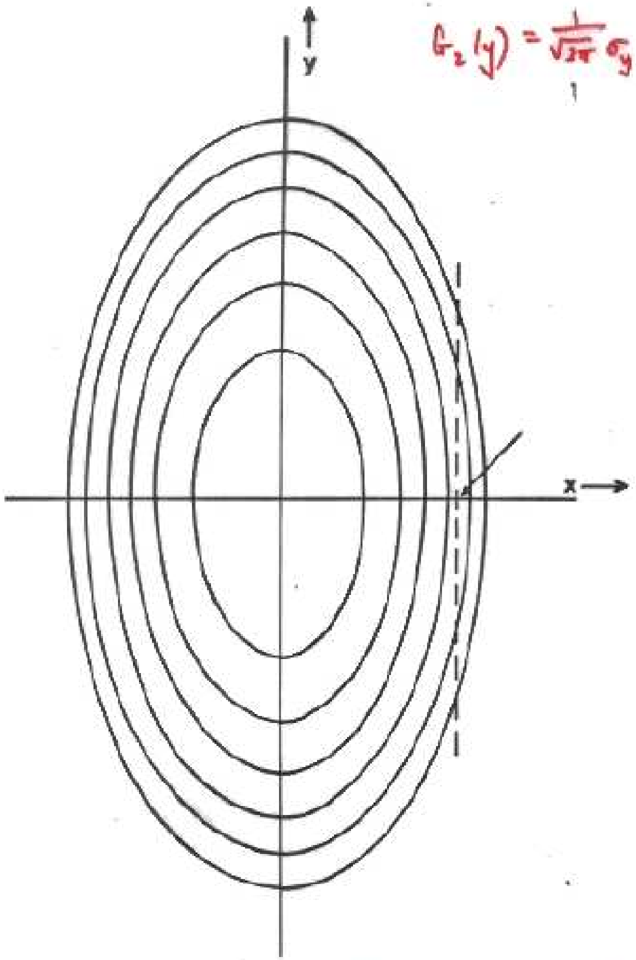
TOWARDS THE ERROR MATRIX

x + y indep Gaussians

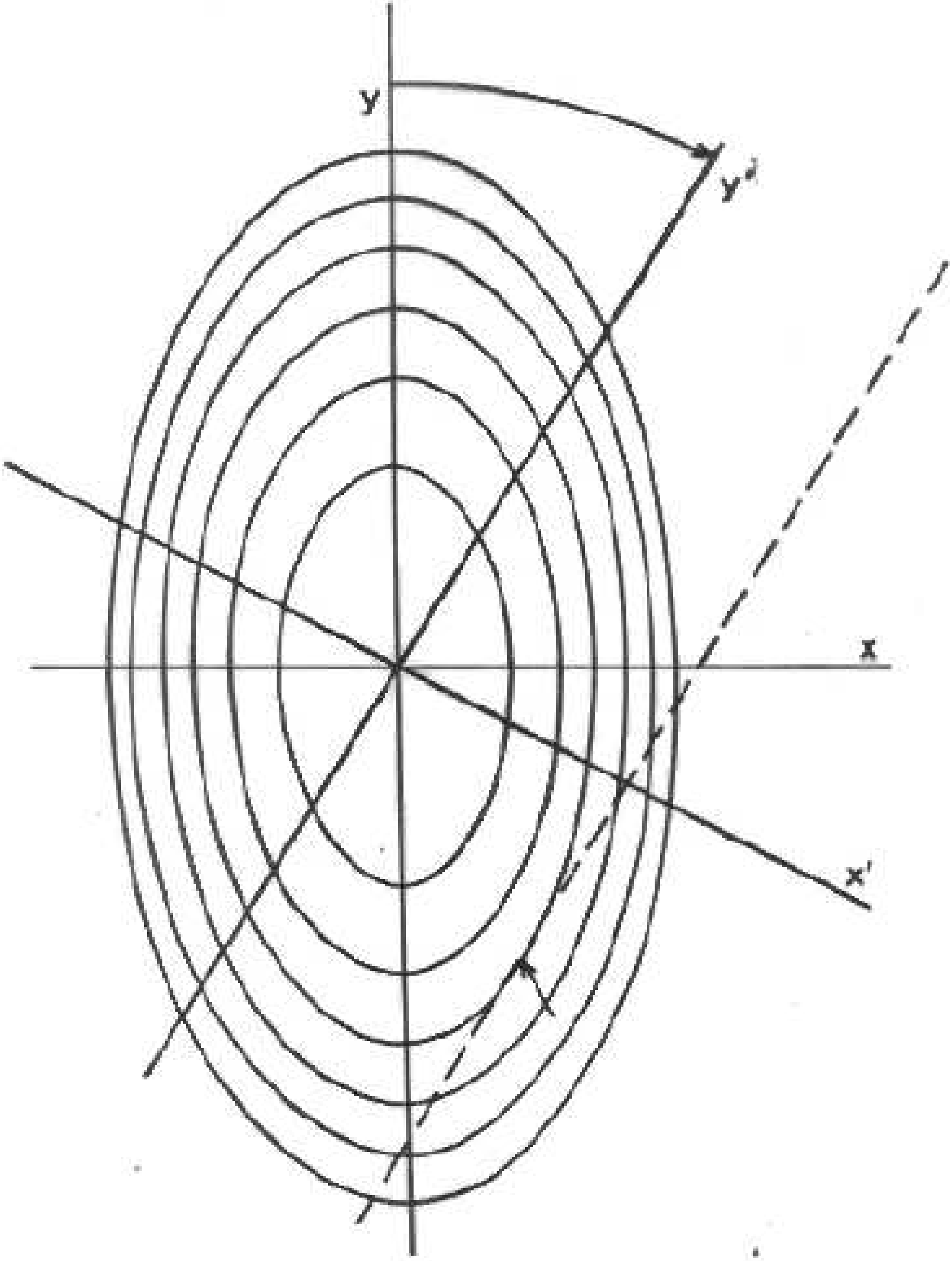
$$P(x,y) = G_1(x)\, G_2(y)$$

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

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



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


y
y'
x
x'

Specific example

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

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

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

Now introduce CORRELATIONS by 30° rotn

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

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

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

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

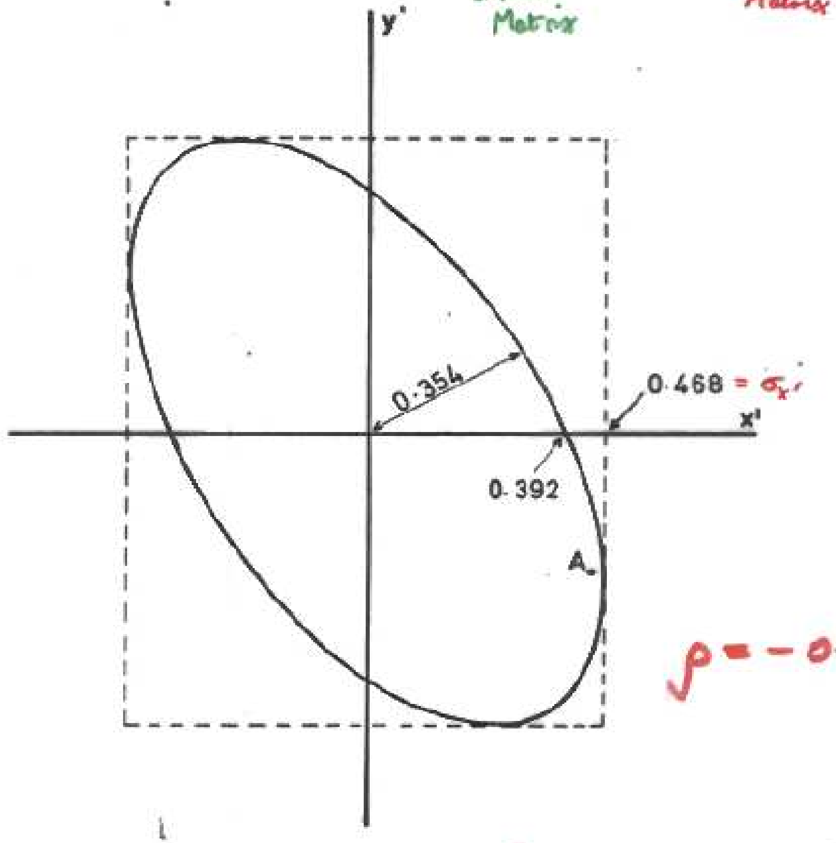
$\frac{1}{2}\left[13x'^2 + 6\sqrt{3}\,x'y' + 7y'^2\right] = 1$

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

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 $= \sigma_x^2$

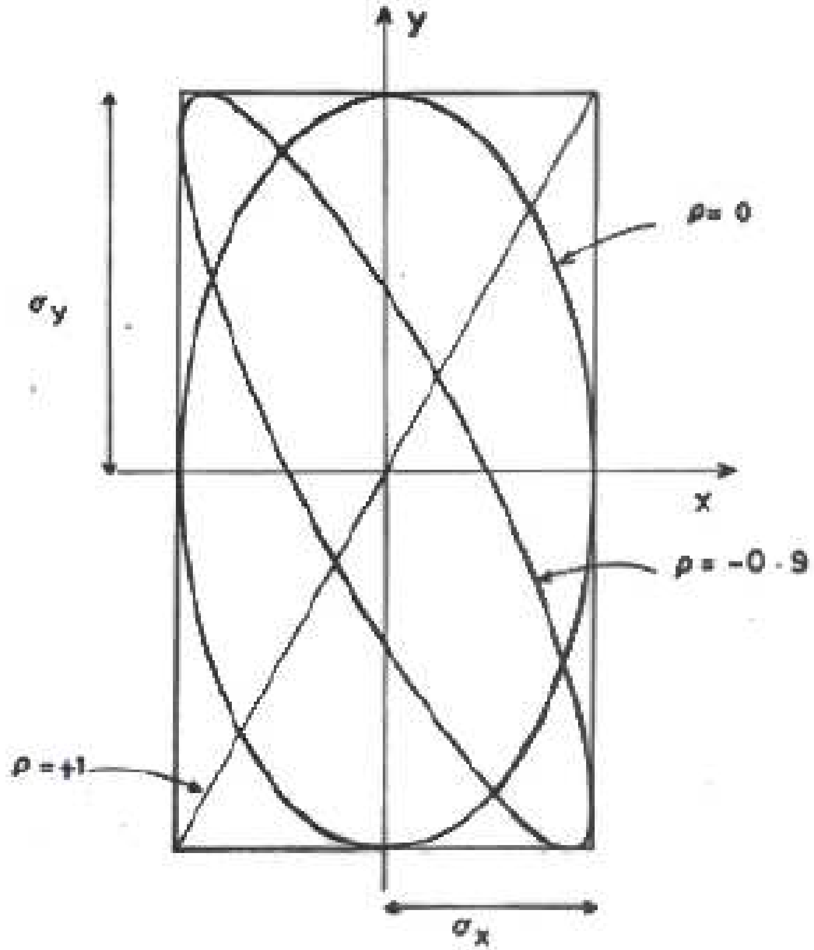
$\left.\begin{matrix}\sigma_x \\ \sigma_y\end{matrix}\right\}$ constant

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 $\qquad x_a = x_a(p_i, p_j)$
$\qquad\qquad\qquad\qquad\qquad\quad x_b = x_b(p_i, p_j)$

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.

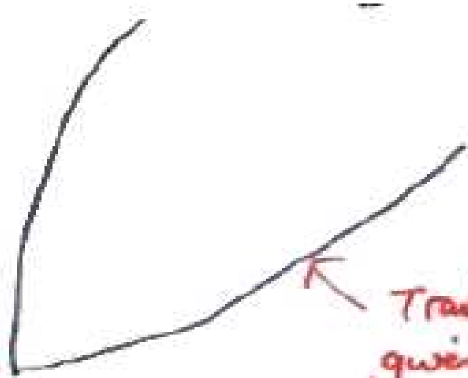
$$\begin{pmatrix} \overline{\delta x_a^2} & \overline{\delta x_a \delta x_b} \\ \overline{\delta x_a \delta x_b} & \overline{\delta x_b^2} \end{pmatrix} = \begin{pmatrix} \frac{\partial x_a}{\partial p_i} & \frac{\partial x_a}{\partial p_j} \\ \frac{\partial x_b}{\partial p_i} & \frac{\partial x_b}{\partial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

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

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

BEWARE!

e.g



Calculate effective mass here

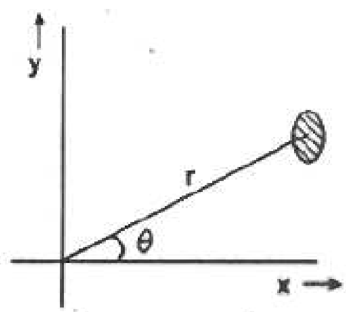
Track params given at centre of track

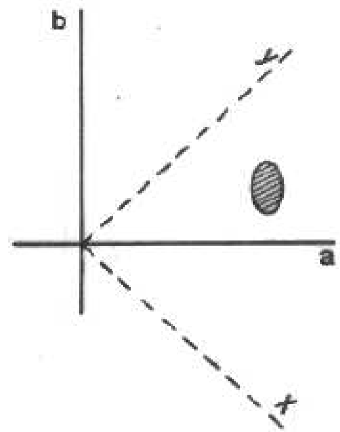
Tracks' error matrix (centre of tracks)

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

Deriv vector for mass in terms of track params at vertex

Transformation matrix from centre of tracks to vertex


y
r
θ
x


b
y
a
x

USING THE ERROR MATRIX

COMBINING RESULTS

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

MORE COMBINING :

SEVERAL PAIRS OF CORRELATED MEAS.

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

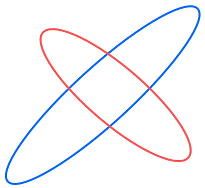
$$S = \sum_i \left\{ (x_i - \hat{x})^2 E^{-1}_{11,i} + (y_i - \hat{y})^2 E^{-1}_{22,i} + 2(x_i - \hat{x})(y_i - \hat{y}) E^{-1}_{12,i} \right\}$$

ice result:—

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

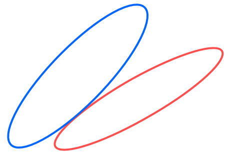
$$= \sum_i \underline{\underline{E}}_i^{-1}$$

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



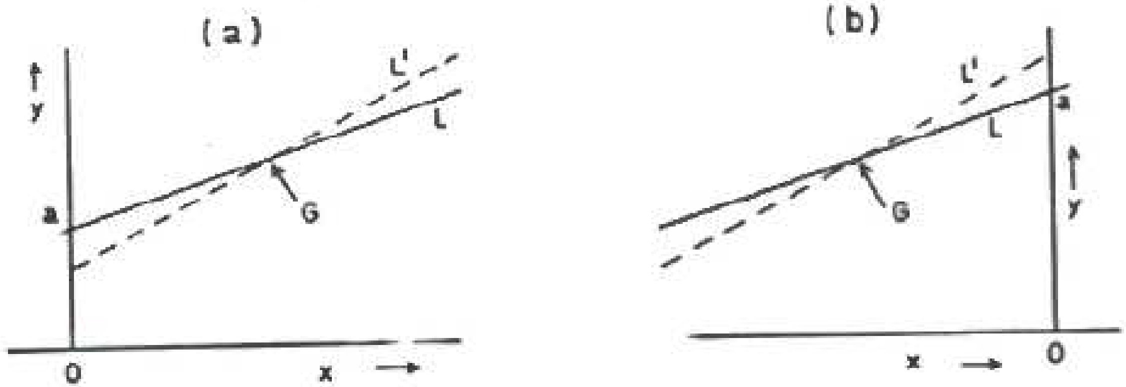
Small error

Example: Chi-sq Lecture

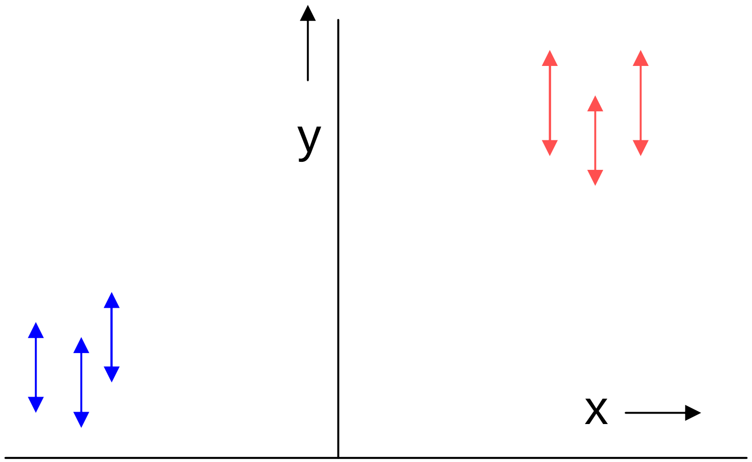


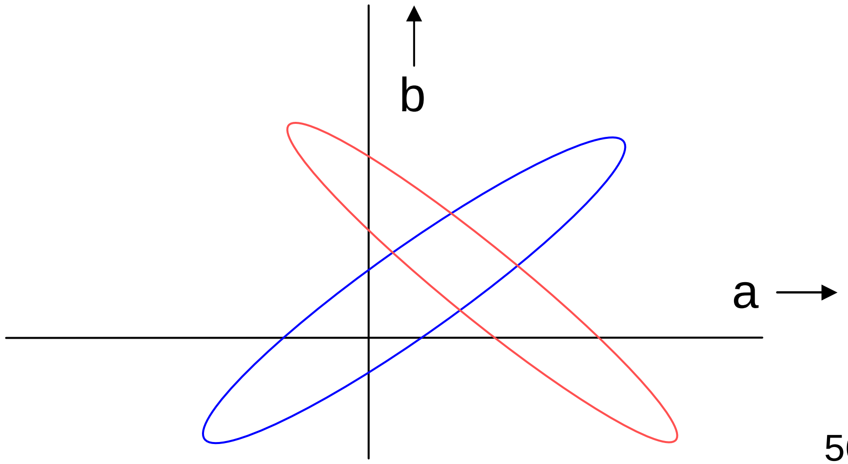
x_{best} outside $x_1 \rightarrow x_2$
y_{best} outside $y_1 \rightarrow y_2$

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

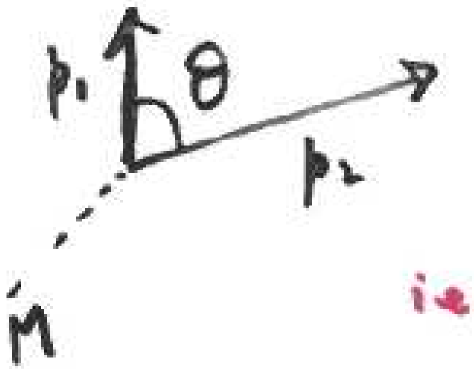


$\langle x \rangle$ pos Fig. 2.4 $\langle x \rangle$ neg





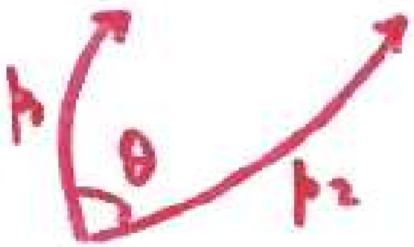
CORRELATIONS + MASS RESOLUTION



$$M^2 = (E_1+E_2)^2 - (\underline{p}_1+\underline{p}_2)^2$$

$$\sim p_1 p_2 \theta \qquad [\, p_i \gg m_i \; ; \; \theta \ll 1 \,]$$

i.e. $M \uparrow$ as $p_i \uparrow$ + $\theta_i \uparrow$



As $p_i \downarrow$, $\theta \uparrow$

$\therefore$ Smaller σ_M

As $p_i \downarrow$, $\theta \downarrow$

$\therefore$ Larger σ_M

ESTIMATING THE ERROR MATRIX

1) ESTIMATE ERRORS

ESTIMATE CORRELATIONS

(Usually easiest if $\rho = 0$ or ± 1)

2) FOR INDEP SOURCES OF ERRORS, ADD ERROR MATRICES

e.g. M_W FROM $WW \to 4$ JETS
$WW \to JJ\ell\nu$

$\underline{\underline{E}}$ = $(M_W)_1, (M_W)_2$ ERROR MATRIX

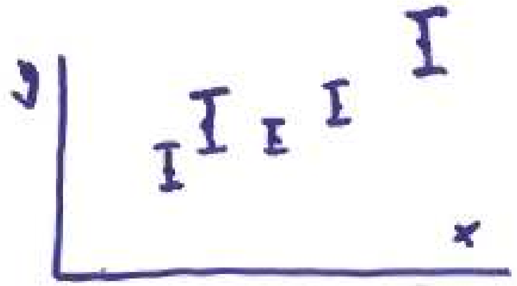
$$\underline{\underline{E}} = \underline{\underline{E}}_{stat} + \underline{\underline{E}}_{B.E.} + \underline{\underline{E}}_{E\ scale} + \underline{\underline{E}}_{FSR} + \underline{\underline{E}}_{colour\ reconn}$$

$$\begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \qquad \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \qquad \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & 0 \end{pmatrix}$$

3) TRANSFORMATIONS

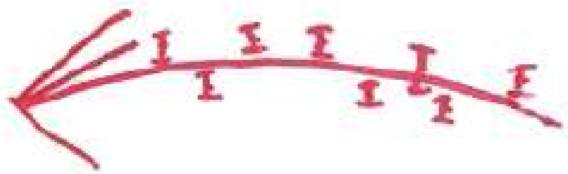
e.g. $(x \pm \sigma_x, \; y \pm \sigma_y)$ with <u>un</u>correl. errors

$\Rightarrow$ r, θ <u>with</u> correlations



Indep data points

$\Rightarrow$ correlated a and b



Track fit

4) REPEATED OBSERVATIONS

$(x_i, y_i) \Rightarrow \sigma_x^2 \quad \sigma_y^2$ and

$\mathrm{cov}(x, y)$ from $\overline{(x-\bar{x})(y-\bar{y})}$

Conclusion

Error matrix formalism makes life easy when correlations are relevant

Next time: $\mathcal{L}$ikelihoods

- What it is
- How it works: Resonance
- Error estimates
- Detailed example: Lifetime
- Several Parameters
- Extended maximum $\mathcal{L}$
- Do's and Dont's with $\mathcal{L}$ ✵✵✵✵