PRACTICAL STATISTICS FOR PHYSICISTS

LOUIS LYONS
CDF, OXFORD

LECTURES AT FERMILAB

August 2004

FNAL LECTURE # 1

INTRODUCTION

LEARNING TO LOVE THE
ERROR MATRIX

STATISTICS FOR NUCLEAR AND PARTICLE PHYSICISTS

L. LYONS

C.U.P (1986)

Available again from C.U.P. ~4 July 99

ERRATA in these lectures

OTHER BOOKS, ETC

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)

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

S L MEYER DATA ANALYSIS FOR SCIENTISTS +
ENGINEERS (Wiley 1975)

A FRODESEN 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

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, $\quad P(A|B) = P(A)$

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

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

$P[\text{Raining} + \text{December}] \neq P(\text{raining}) \times \frac{1}{12}$ $\quad$ ~~INDEP~~

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

~~INDEP~~

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

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

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

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

PROBABILITY | STATISTICS

Example : Dice

PROBABILITY	STATISTICS
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	
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?
	PARAM DETERM.
If unbiassed, what is P(n evens out of 100 trials)?	Observe 65 evens in 100 trials. Is it unbiassed?
	GOODNESS OF FIT?
	Or is $P(\text{even}) = \frac{2}{3}$?
	HYPOTHESIS TESTING
THEORY $\Rightarrow$ DATA	DATA $\Rightarrow$ THEORY

N.B. PARAM DETERMINATION not sensible if GOODNESS OF FIT is poor/bad

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 EXPERIMENTS

(uncorrelated)

$x_i \pm \sigma_i$

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

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

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

Minimise S

σ from $S_{min} + 1$

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

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

$$\left.\begin{matrix} 100 \pm 10 \\ 1 \pm 1 \end{matrix}\right\} \Longrightarrow \begin{matrix} 2 \pm 1 \;? \\ \text{or } 50.5 \pm 5 \;? \end{matrix}$$

DIFFERENCE BETWEEN ADDING + AVERAGING

NO OF MARRIED MEN = 10.0 ± 0.5 Million

NO OF MARRIED WOMEN = 8 ± 3 Million

Total = 18 ± 3 million

Average = 9.9 ± 0.5

$\Rightarrow$ total = 20 ± 1 million

General point: Including theoretical input can improve accuracy of answer

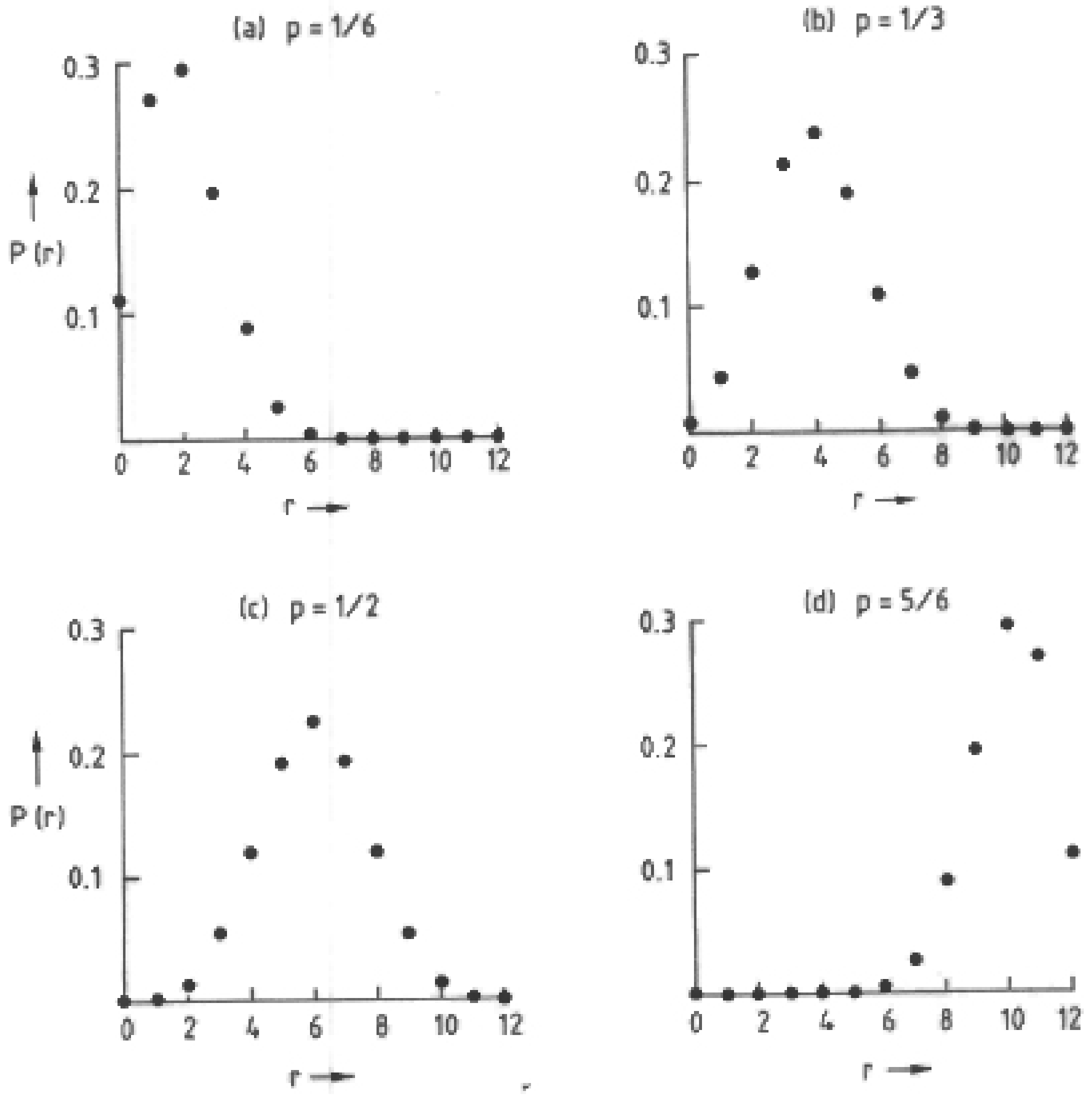


Fig. A3.1 The probabilities $P(r)$, according to the binomial distribution, for r successes out of 12 independent trials, when the probability p of success in an individual trial is as specified in the diagram. As the expected number of successes is $12p$, the peak of the distribution moves to the right 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.

Thus the expected number of successes of our die-throwing experiment was $12 \times (1/6) = 2$, with a variance of $12 \times (1/6) \times (5/6) = 5/3$ (or a standard deviation of $\sqrt{5/3}$). This tells us that we cannot expect that the number of successes will be much larger than a couple of times $\sqrt{5/3}$ above 2, i.e. more than five 6's is unlikely (see Fig. A3.1(a)).

For the same experiment of throwing a die 12 times, we could have

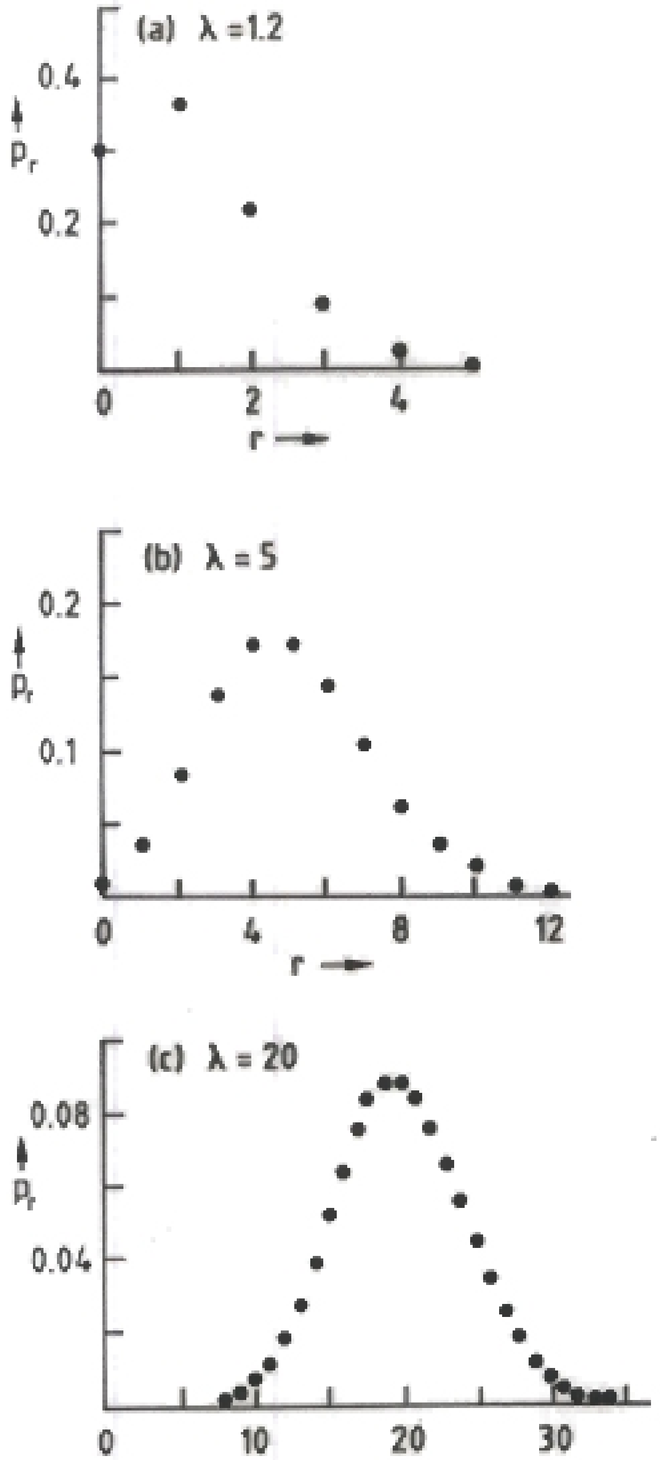
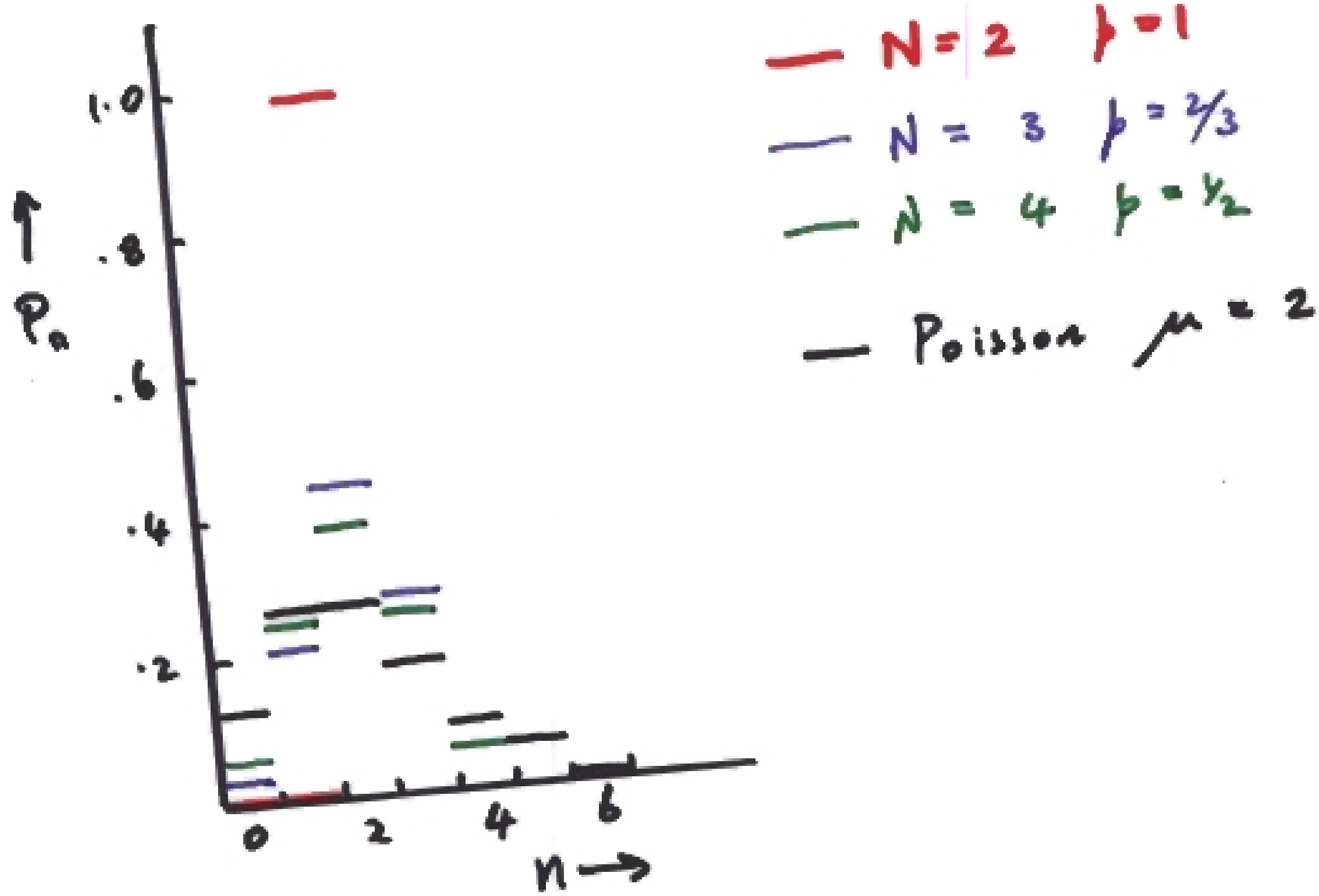


Fig. A4.1 Poisson distributions for different values of the parameter λ. (a) $\lambda = 1.2$; (b) $\lambda = 5.0$; (c) $\lambda = 20.0$. P_r is the probability of observing r events. (Note the different scales on the three figures.) For each value of λ, the mean of the distribution is at λ, and the RMS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

In a similar way, the Poisson distribution is likely to be applicable to

BINOMIAL ⟹ POISSON



RELATION BETWEEN POISSON AND BINOMIAL

N people at lecture, m males & f females

Assume that these are representative of basic rates :–

ν people $\qquad\qquad$ νp males $\qquad$ $\nu(1-p)$ females

Probability of observing N people

$$P_{\text{Poisson}} = \frac{e^{-\nu}\,\nu^N}{N!}$$

Probability of given male/female division

$$P_{\text{Binomial}} = \frac{N!}{m!\,f!}\,p^m(1-p)^f$$

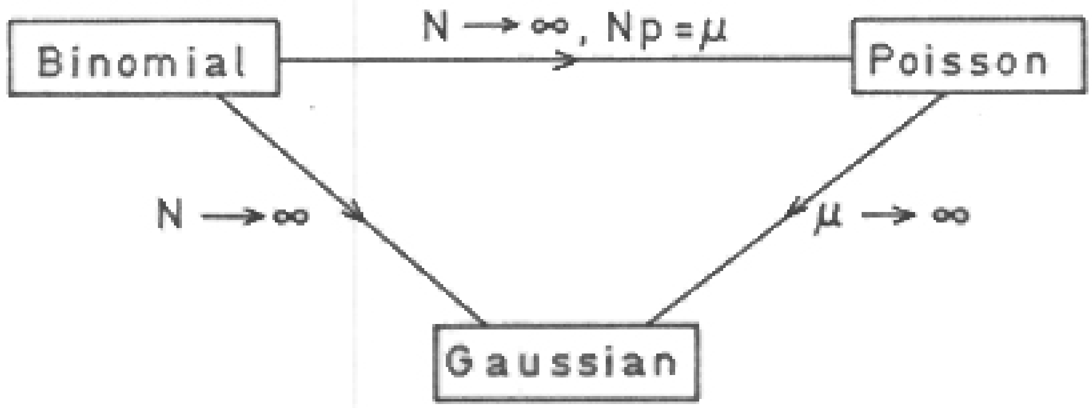
Probability of N people, m males & f females

$$P = P_{\text{Poisson}}\,P_{\text{Binomial}}$$

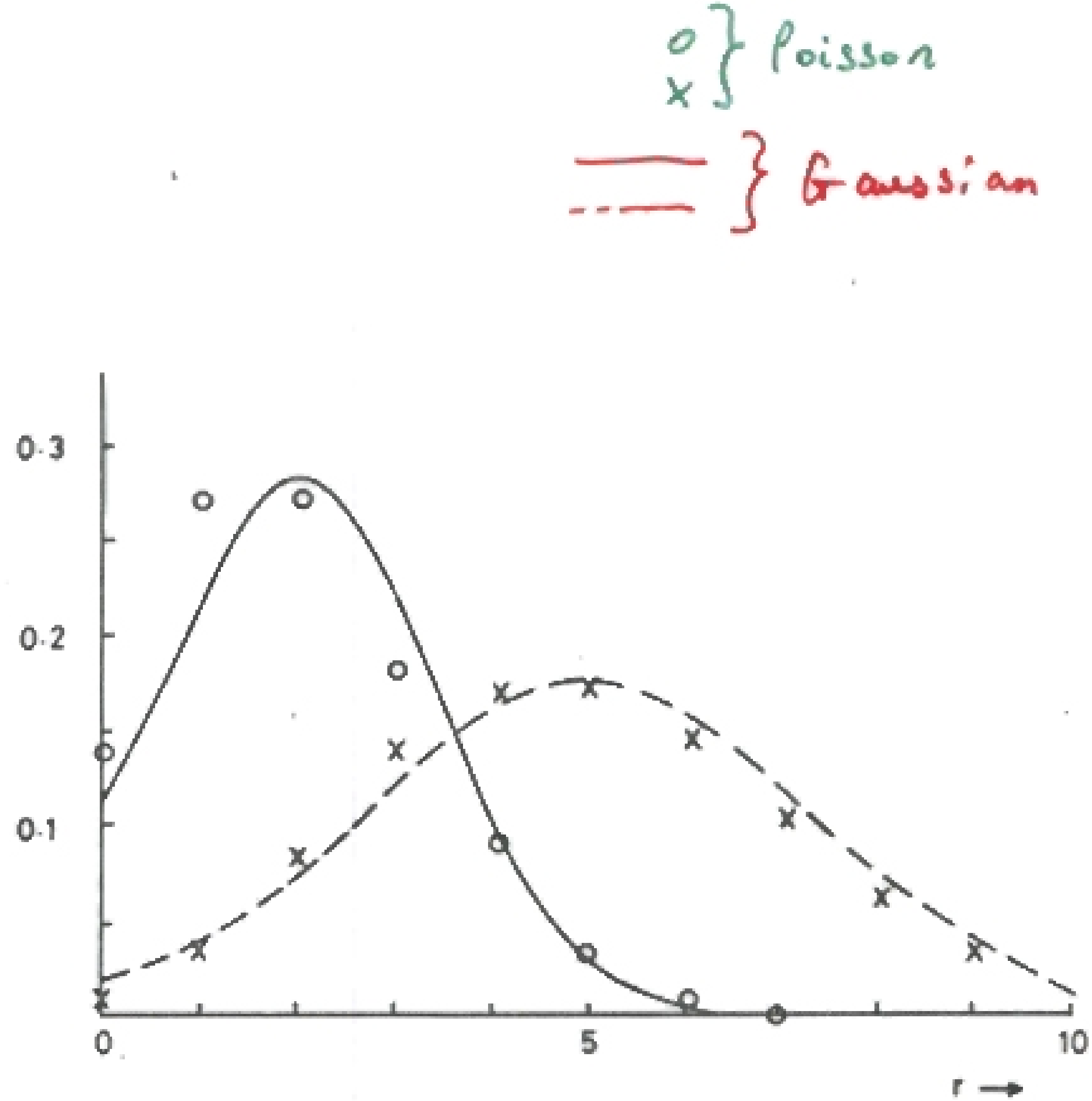
$$= \left\{\frac{e^{-\nu p}\,\nu^m p^m}{m!}\right\} \times \left\{\frac{e^{-\nu(1-p)}\,\nu^f(1-p)^f}{f!}\right\}$$

= Poisson distribution for males × Poisson distribution for females

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



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



Relevant for Hypothesis Testing

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

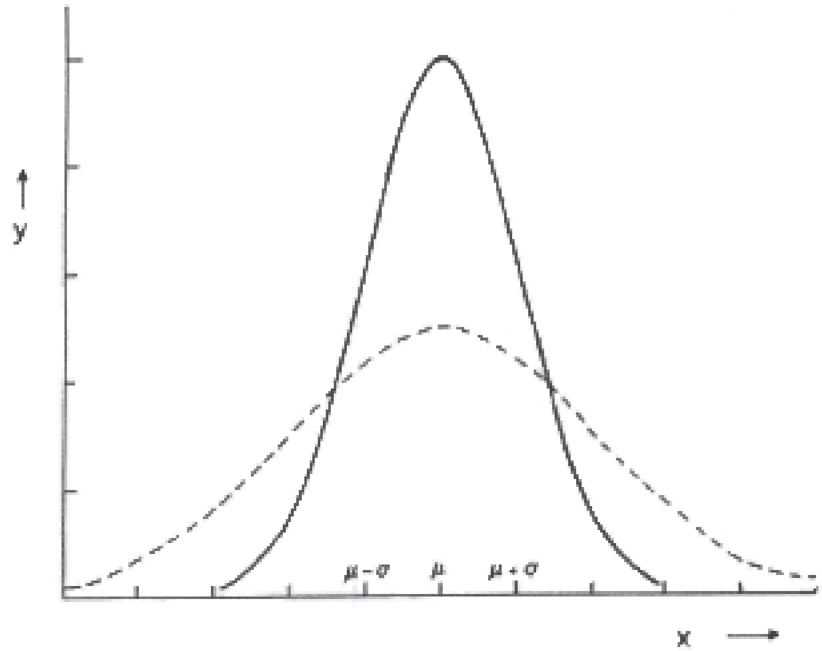


Fig. 1.5. The solid curve is the Gaussian distribution of eqn (1.14). The distribution peaks at the mean μ, and its width is characterised by the parameter σ. The dashed curve is another Gaussian distribution with the same values of μ, but with σ twice as large as the solid curve. Because the normalisation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

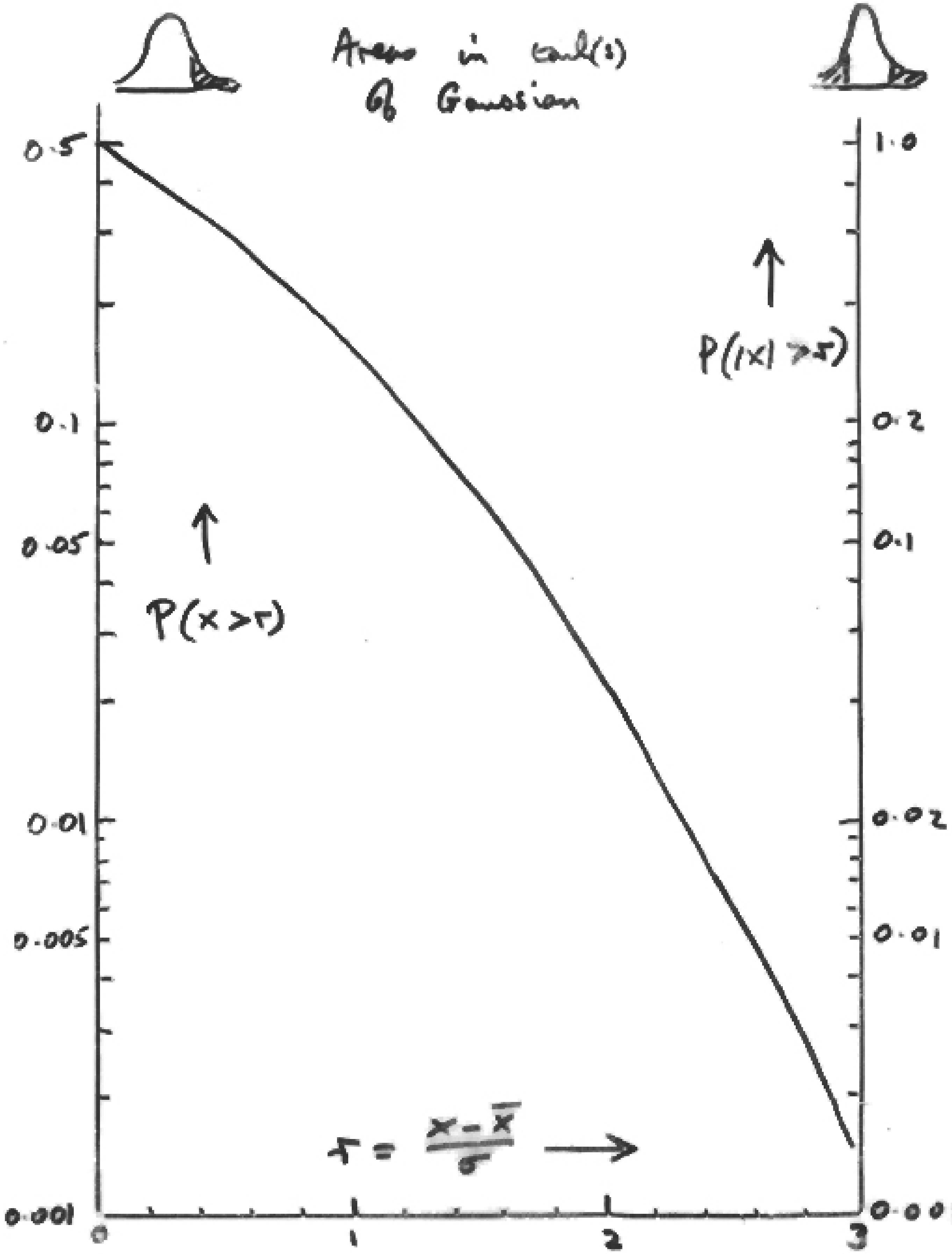
i) RMS of Gaussian $= \sigma$
(Hence factor of 2 in defn of Gaussian)

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



STUDENT'S t

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

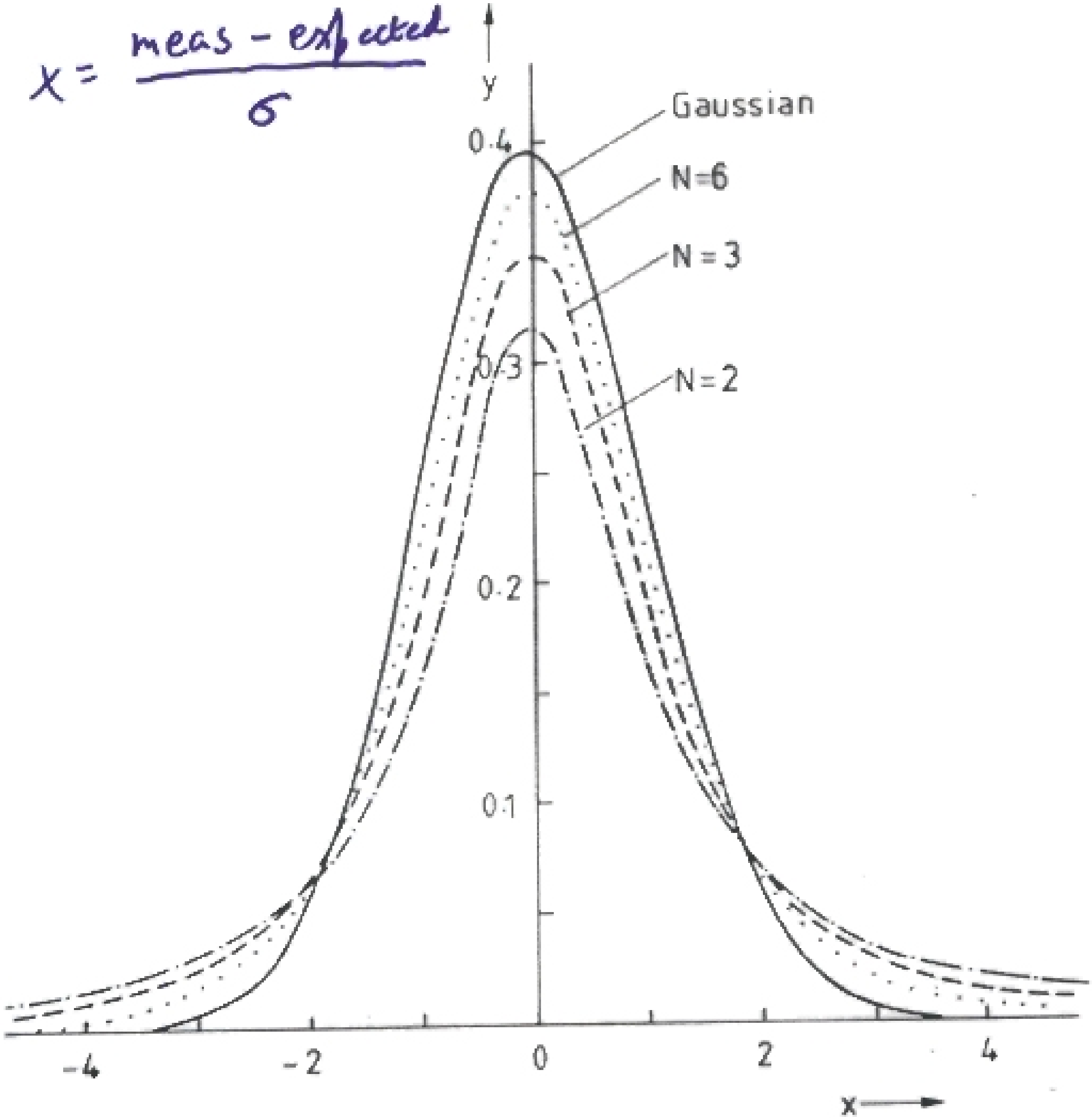


Fig. A5.1 Comparison of Student's t distributions for various values of the number of observations N, with the Gaussian distribution, which is the limit of the Student's distributions as N tends to infinity.

STUDENT'S T

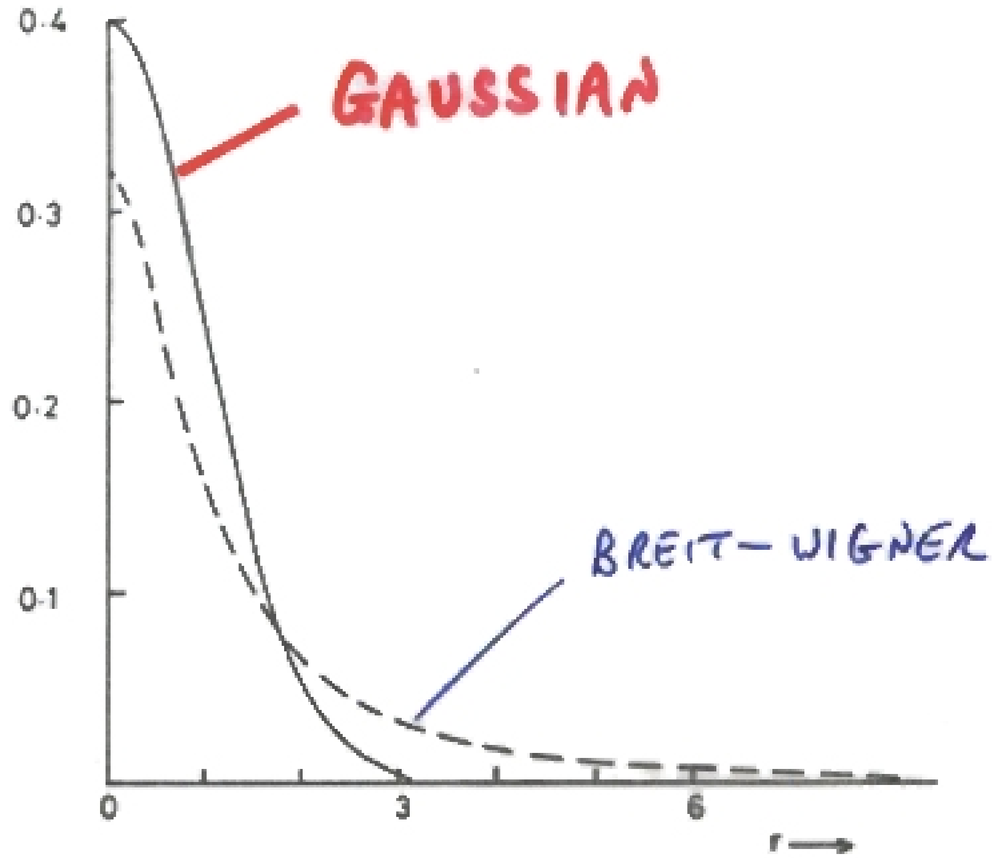
PROB $(t > t_0)$

NDF	0.5%	1%	2.5%	5%	10%	15%
1	64	32	13	6	3.1	1.96
2	10	7	4.3	2.9	1.89	1.37
5	4.0	3.4	2.6	2.0	1.48	1.16
10	3.2	2.8	2.2	1.81	1.37	1.10
30	2.8	2.5	2.0	1.70	1.31	1.06
∞	2.6	2.3	1.96	1.64	1.28	1.04

$$t = \frac{\bar{x} - \mu}{s}$$

Prob $(|t| > t_0) = 2 *$ top line

[NDF $= \infty$] is equivalent to Gaussian.



$$\text{Gaussian} = N(0,1)$$

$$\text{B-W} = \frac{1}{\pi}\frac{1}{r^2+1}$$

Gaussian in 2-variables

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

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

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} is $\langle (x_i - \bar{x}_i)(x_j - \bar{x}_j) \rangle$]

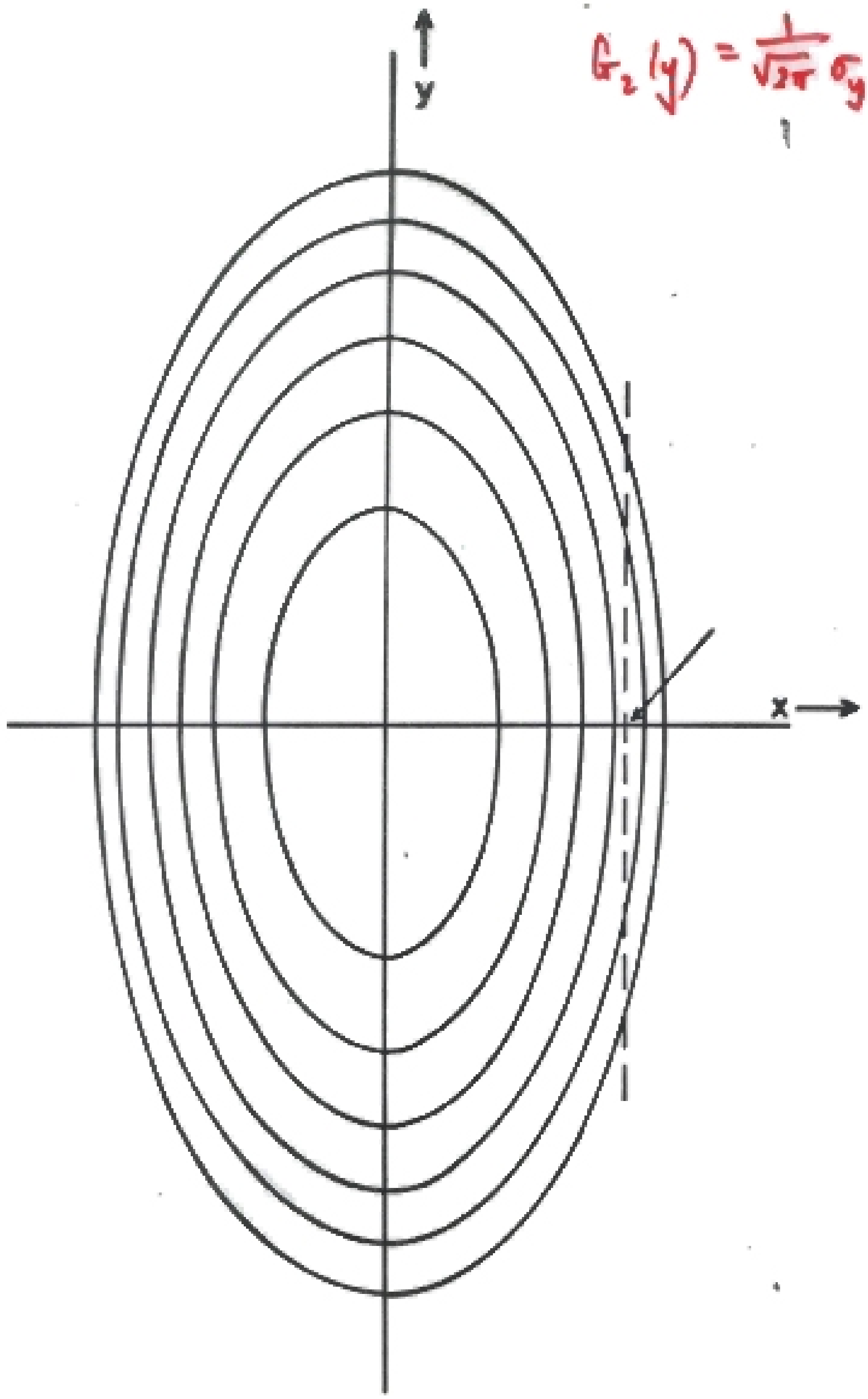
TOWARDS THE ERROR MATRIX

$x + y$ indep Gaussians

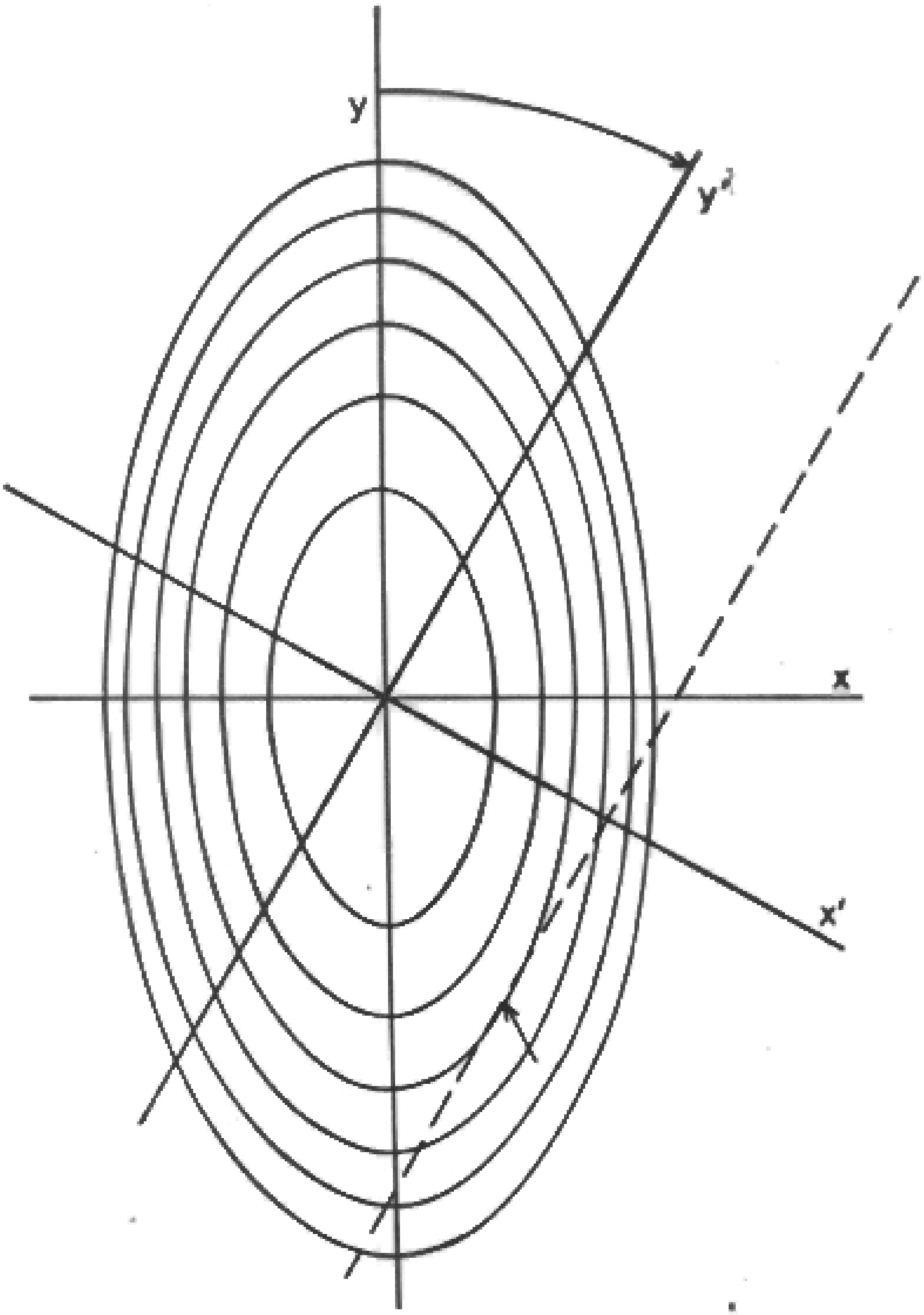
$$P(x,y) = G_1(x)\, G_2(y)$$

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

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



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


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

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

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

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

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

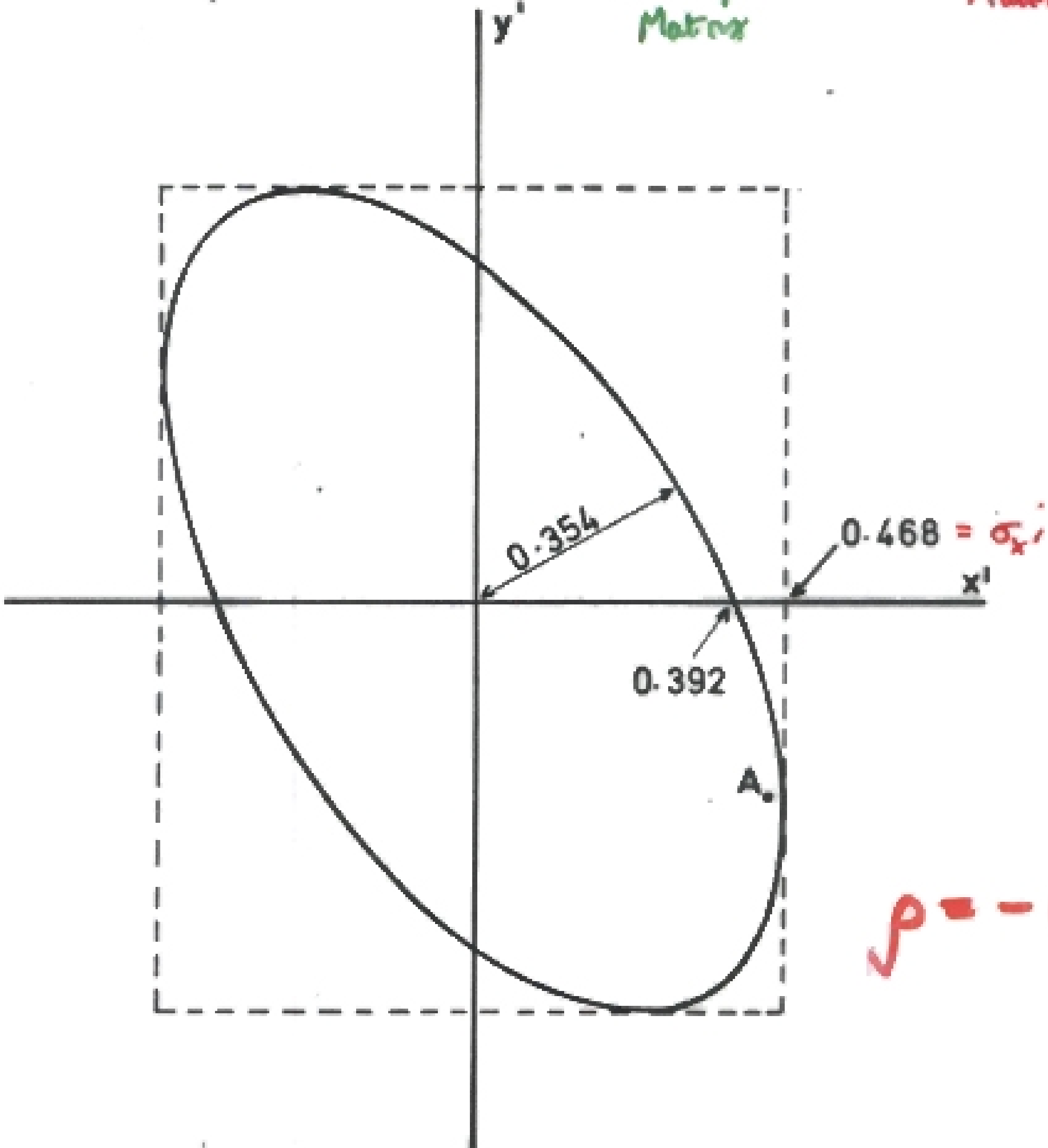
$\frac{1}{2}\left[13x'^2 + 6\sqrt{3}\,x'y' + 7y'^2\right] = 1$

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

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 $= \sigma_x^2$

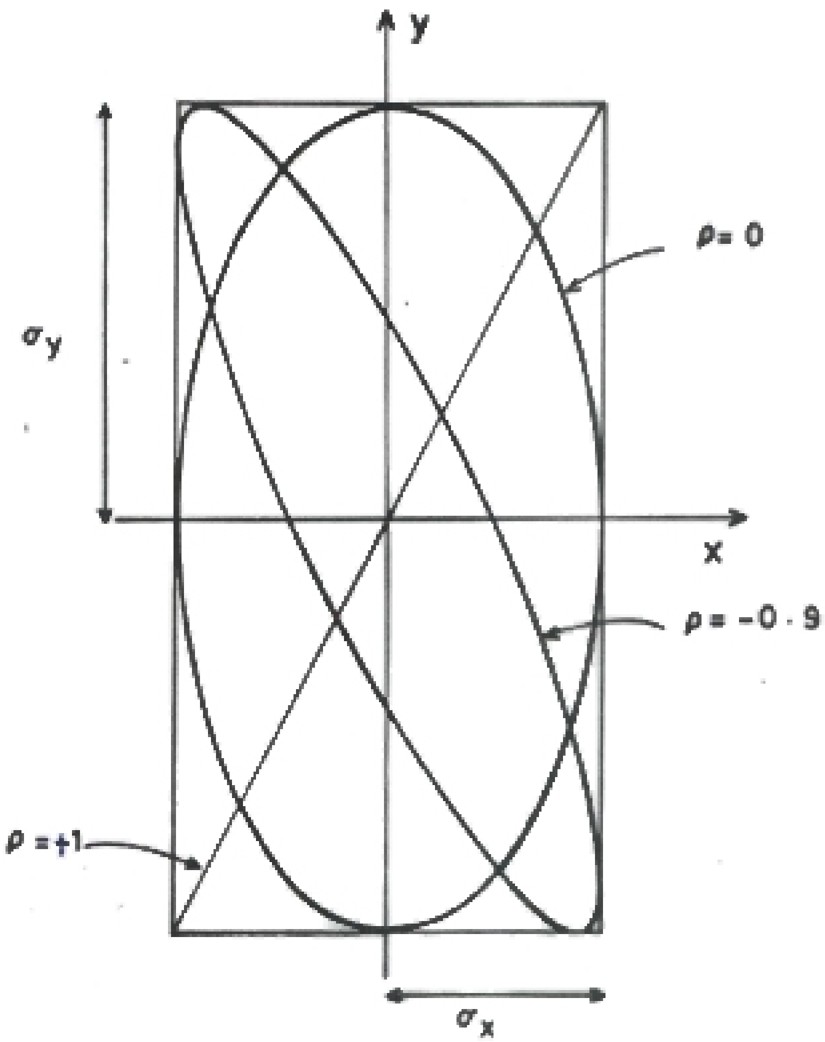
$\left.\begin{matrix}\sigma_x \\ \sigma_y\end{matrix}\right\}$ constant

ρ varying

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

Covariance

Error Matrix



18

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

$$\overline{\delta y^2} = \underbrace{\begin{pmatrix} \frac{\partial y}{\partial x_a} & \frac{\partial y}{\partial x_b} \end{pmatrix}}_{\tilde{D}} \underbrace{\begin{pmatrix} \overline{\delta x_a^2} & \overline{\delta x_a \delta x_b} \\ \overline{\delta x_a \delta x_b} & \overline{\delta x_b^2} \end{pmatrix}}_{\text{Error matrix}} \underbrace{\begin{pmatrix} \frac{\partial y}{\partial x_a} \\ \frac{\partial y}{\partial x_b} \end{pmatrix}}_{\text{Derivative vector } D}$$

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

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

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

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!

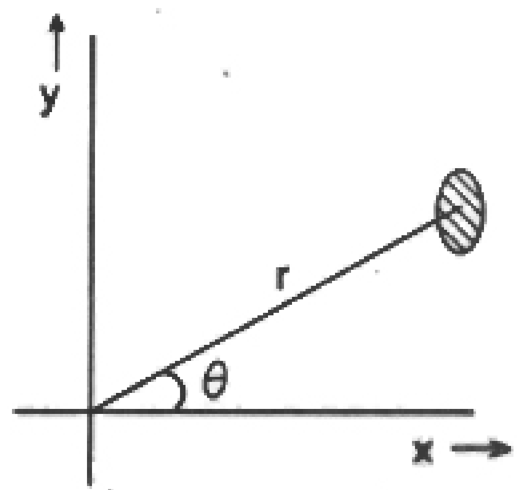
e.g.

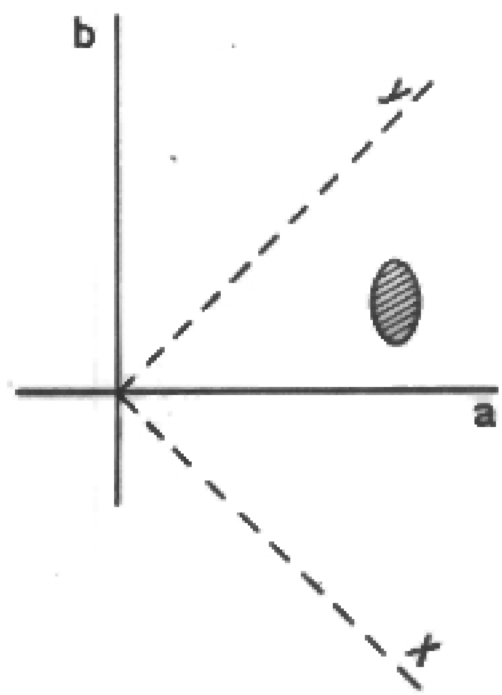
Tracks' error matrix (centre of tracks)

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

Transformation matrix from centre of tracks to vertex

Deriv vector for mass in terms of track params at 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

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

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

FOR UNCORRELATED

MORE COMBINING :

SEVERAL PAIRS OF CORRELATED MEAS.

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

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

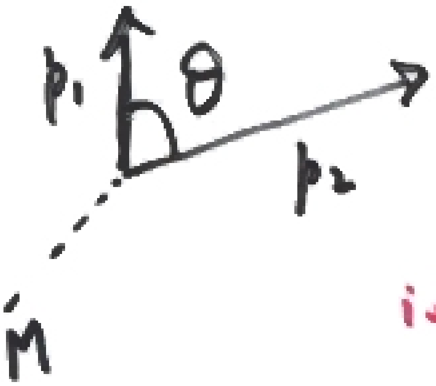
ice result:—

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.

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

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

$$\underline{\underline{E}}_{stat} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \qquad \underline{\underline{E}}_{B.E.} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \qquad \underline{\underline{E}}_{colour\ +recon} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & 0 \end{pmatrix}$$

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 $\overline{(x-\bar{x})(y-\bar{y})}$