Practical Statistics for Physicists

Louis Lyons

Imperial College and Oxford

CMS expt at LHC

l.lyons@physics.ox.ac.uk

CERN Latin American School

March 2015

Books

Statistics for Nuclear and Particle Physicists

Cambridge University Press, 1986

Available from CUP

Errata in these lectures

Topics

1) Introduction

2) Bayes and Frequentism

3) χ^2 and $\mathcal{L}$ikelihoods

4) Higgs: Example of Search for New Physics

Time for discussion

Introductory remarks

What is Statistics?

Probability and Statistics

Why uncertainties?

Random and systematic uncertainties

Combining uncertainties

Combining experiments

Binomial, Poisson and Gaussian distributions

What do we do with Statistics?

Parameter Determination (best value and range)

e.g. Mass of Higgs = 80 $\pm$ 2

Goodness of Fit

Does data agree with our theory?

Hypothesis Testing

Does data prefer Theory 1 to Theory 2?

(Decision Making

What experiment shall I do next?)

Why bother?

HEP is expensive and time-consuming

so

Worth investing effort in statistical analysis

→ better information from data

Probability and Statistics

Example: Dice

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

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

THEORY→ DATA

Given 20 5's in 100 trials, what is P(5)? And its unceretainty?

Given 60 evens in 100 trials, is it unbiassed?

Or is P(evens) =2/3?

DATA→ THEORY

Probability and Statistics

Example: Dice

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

Given 20 5's in 100 trials, what is P(5)?
And its uncertainty?
Parameter Determination

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

Given 60 evens in 100 trials, is it unbiassed?
Goodness of Fit

Or is P(evens) =2/3?
Hypothesis Testing

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

Why do we need uncertainties?

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 Uncertainties

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 uncertainties: 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 uncertainties 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.

Combining uncertainties

$z = x - y$

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

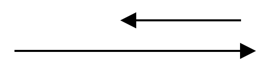
Why $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$? [2]

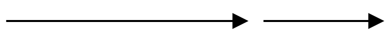
Combining uncertainties

$$z = x - y$$

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

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……………

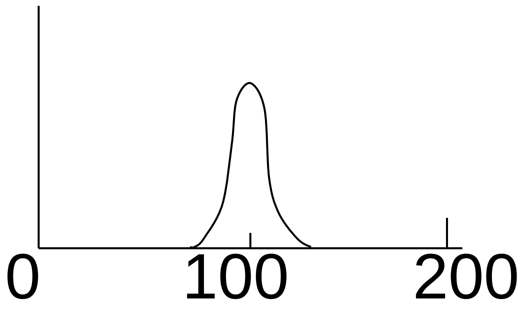
3) **Averaging is good for you**: N measurements $x_i \pm \sigma$

[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

(Prob ~ 1% for 100 people doing it 1/μsec since Big Bang)

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, $x/\sqrt{y}$,…….

3) Anything else:

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

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

OR numerically:

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

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

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

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

N.B. All formulae approx (except 1)) – assumes small uncertainties

COMBINING EXPERIMENTS

$x_i \pm \sigma_i$ (uncorrelated)

$\hat{x} = \frac{\sum x_i/\sigma_i^2}{\sum 1/\sigma_i^2}$ ← Minimise S — From $S = \sum (x_i - \hat{x})^2/\sigma_i^2$

$1/\sigma^2 = \sum 1/\sigma_i^2$ ← σ from $S_{min} + 1$

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

Define $w_i = 1/\sigma_i^2$ = weight ~ information content

$\hat{x} = \sum w_i x_i / \sum w_i$

$W = \sum w_i$

N.B. Better to combine data!

Example: Equal σ_i $\Rightarrow$ $\hat{x} = \bar{x}$

$\sigma = \sigma_i/\sqrt{n}$

BEWARE

100±10
1±1

2±1?

or 50.5±5?

Difference between averaging and adding

Isolated island with conservative inhabitants

How many married people ?

Number of married men = 100 ± 5 K

Number of married women = 80 ± 30 K

Total = 180 ± 30 K

Wtd average = 99 ± 5 K

Total = 198 ± 10 K

CONTRAST

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 cosθ? Prob of 52/70 events with cosθ > 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 = $\Sigma s P_s$ = Np,

as is obvious

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

Variance ~ Np, for p~0

~ N(1-p) for p~1

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

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→ ∞: Binomial → Gaussian

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

- N→ ∞, p→0, Np = const: Binomial → Poisson

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

{N.B. Gaussian continuous and extends to -∞}

Binomial Distributions

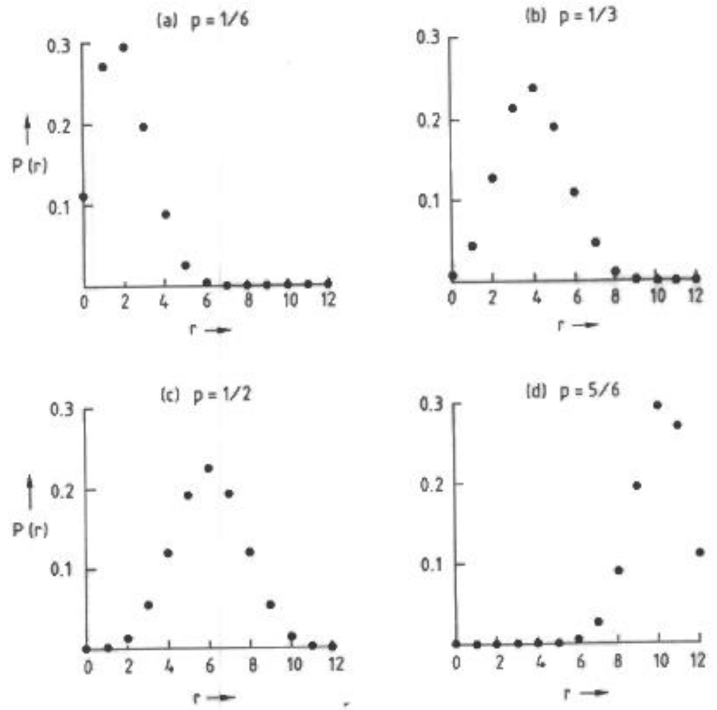


Fig. A3.1 The probabilities $P(r)$, according to the binomial distribution, for r successes out of 12 independent trials, when the probability p of success in an individual trial is as specified in the diagram. As the expected number of successes is $12p$, the peak of the distribution moves to the right 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 $P_n = e^{-\mu} \mu^n/n!$

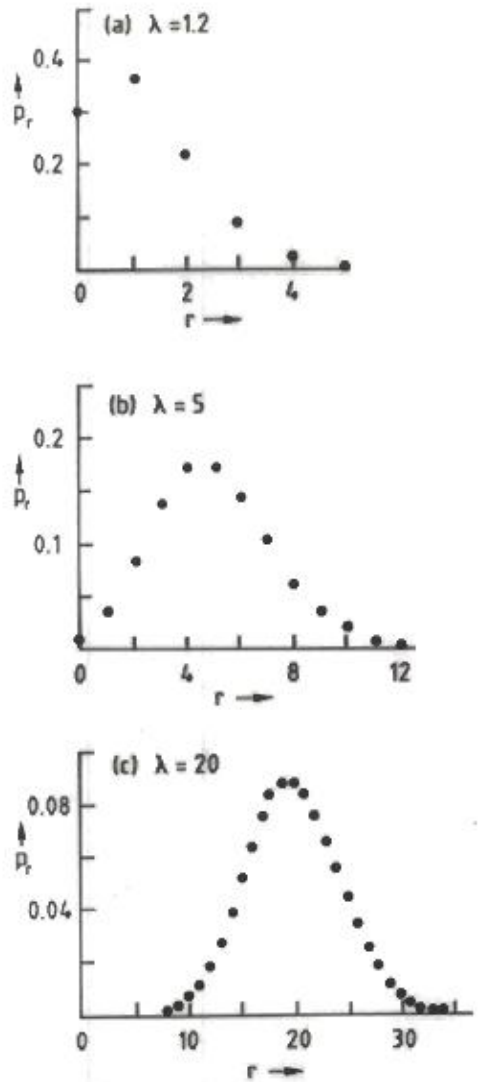
$P_0 = e^{-\mu}$ $P_1 = \mu\, e^{-\mu}$ $P_2 = \mu^2/2\ e^{-\mu}$

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 $\sim \mu^2$?

Poisson Distributions



Approximately Gaussian

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.

Gaussian or Normal

14 *Experimental errors*

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

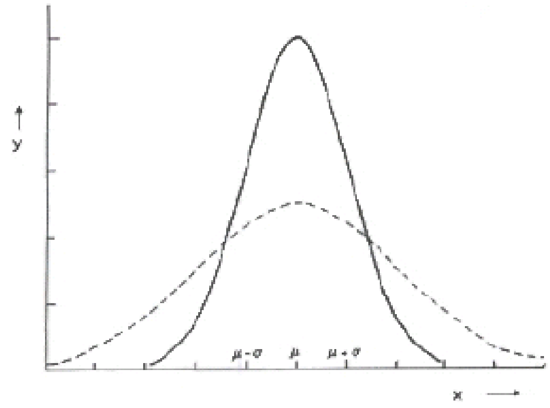
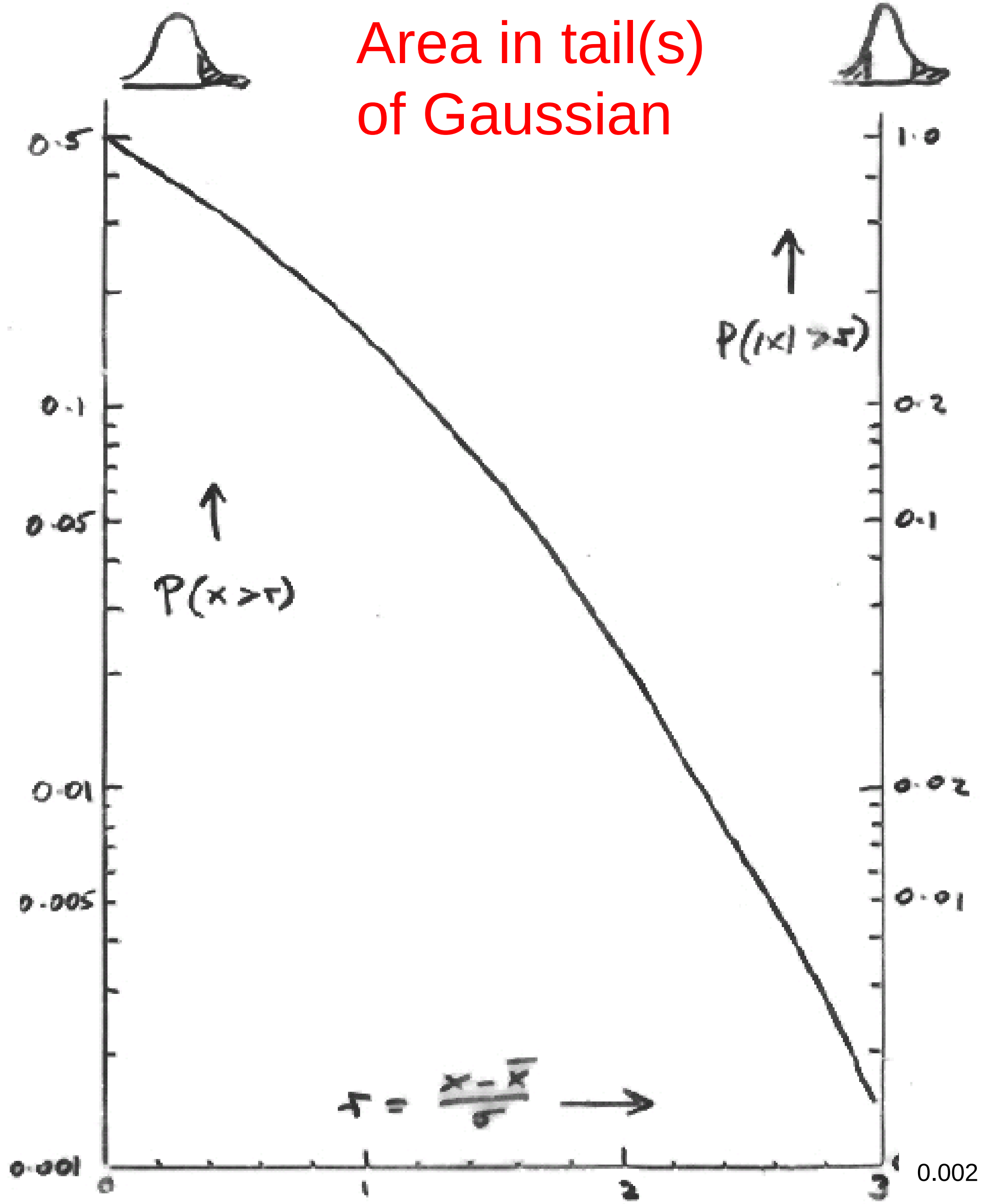


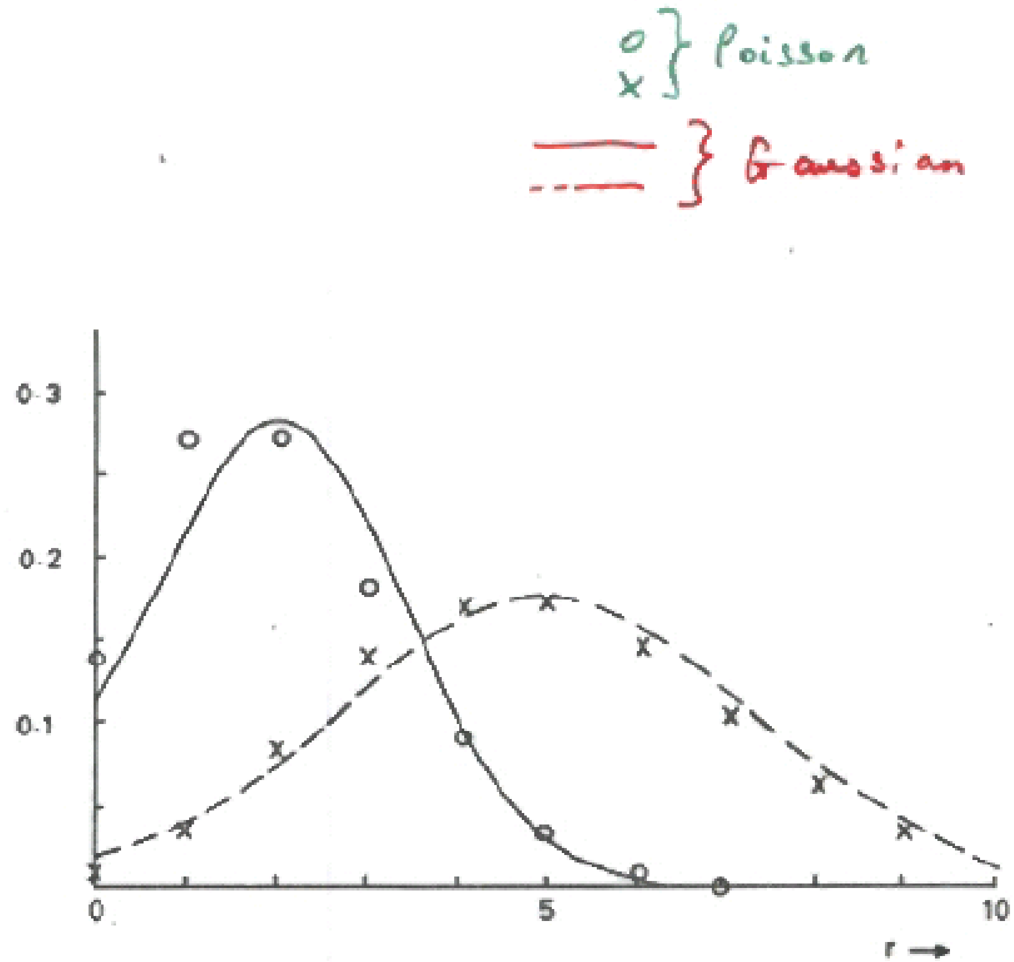
Fig. 1.5. The solid curve is the Gaussian distribution of eqn (1.14). The distribution peaks at the mean μ, and its width is characterised by the parameter σ. The dashed curve is another Gaussian distribution with the same values of μ, but with σ twice as large as the solid curve. Because the normalisation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

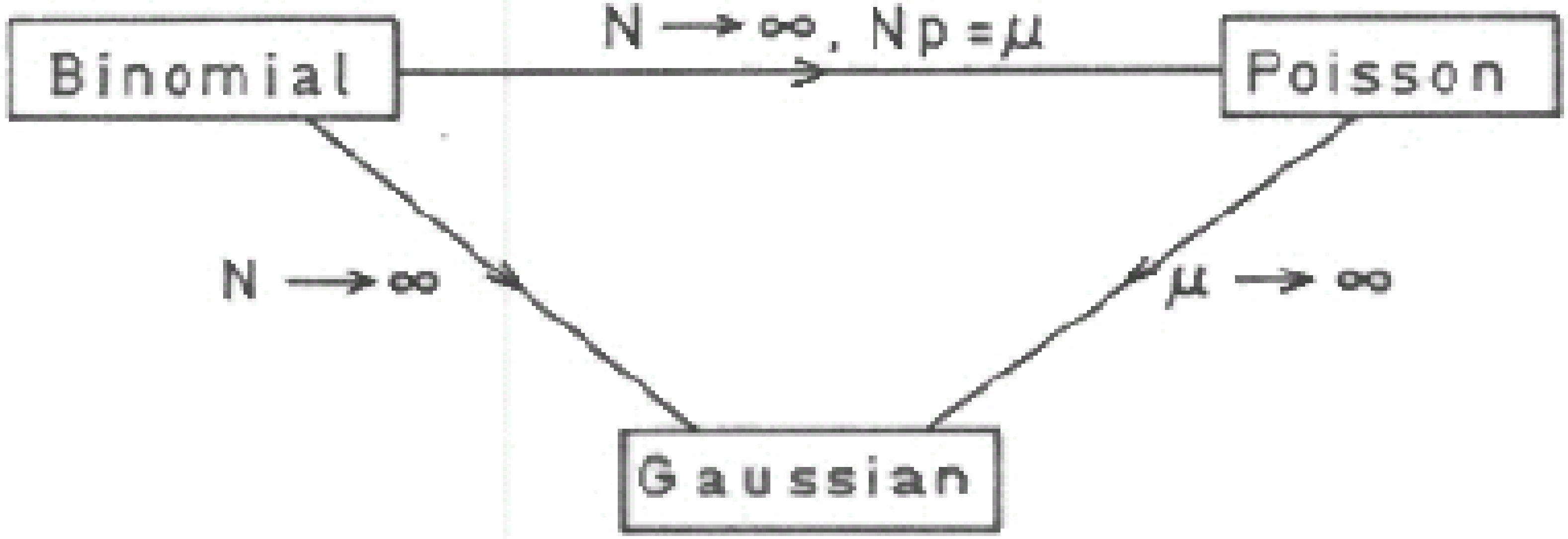
i) RMS of Gaussian = σ
(hence factor of 2 in definition of Gaussian)
ii) At $x = \mu \pm \sigma$, $y = y_{max}/\sqrt{e}$ ~0.606 y_{max}
(i.e. σ = half-width at 'half'-height)
iii) Fractional area within $\mu \pm \sigma$ = 68%
iv) Height at max = $1/(\sigma\sqrt{2\pi})$

Area in tail(s) of Gaussian





Relevant for Goodness of Fit


Binomial
N → ∞, Np = μ
Poisson
N → ∞
μ → ∞
Gaussian