

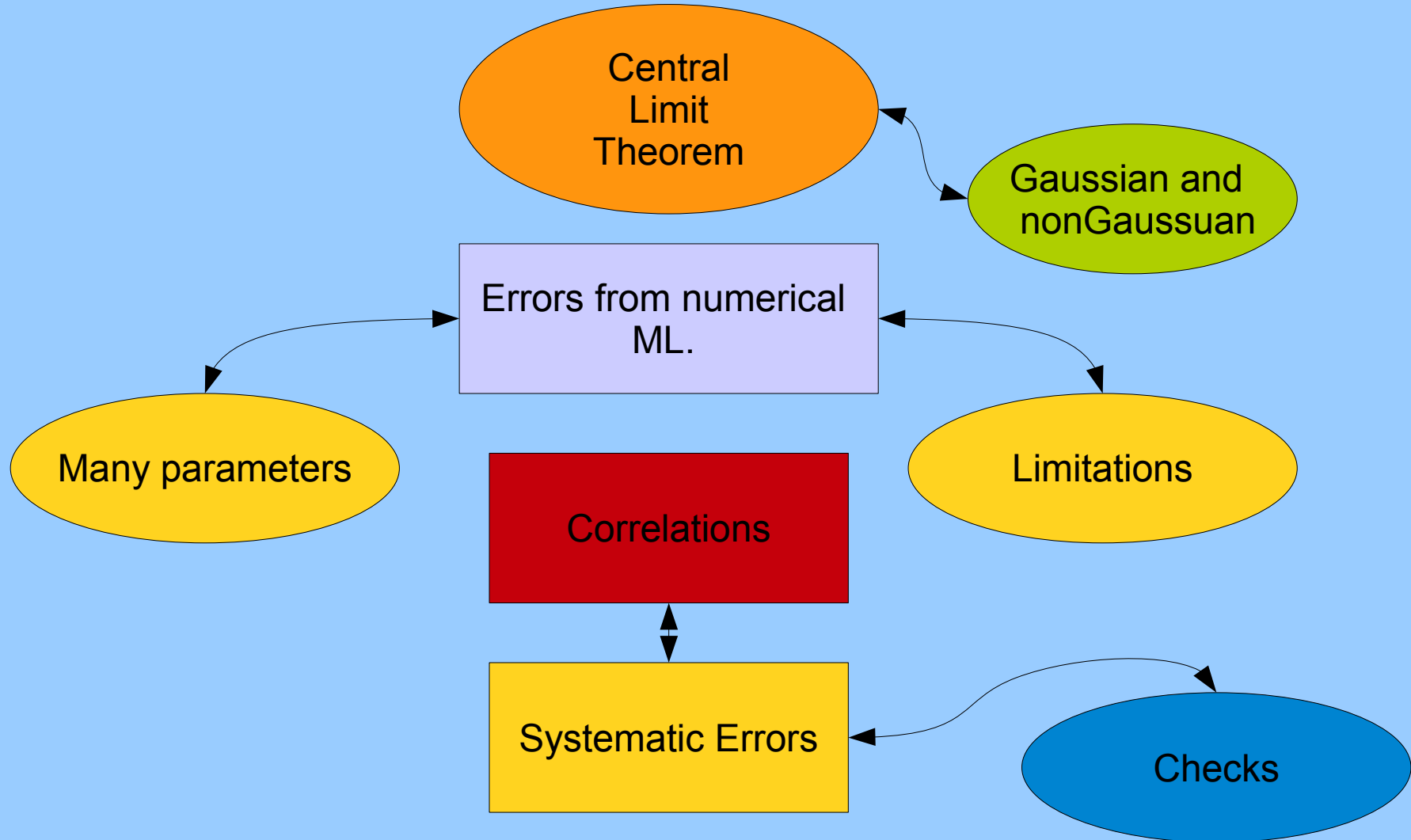
# Statistics (3)

## Errors

*Roger Barlow*  
*Manchester University*

IDPASC school  
Sesimbra  
15<sup>th</sup> December 2010

# Summary



# The Central Limit Theorem

Suppose a random variable  $x$  is the sum of several independent identically (or similarly) distributed variables  $x_1, x_2, x_3, \dots, x_N$ . Then

(1) The mean of the distribution for  $x$  is the sum of the means:  $\mu = \mu_1 + \mu_2 + \mu_3 + \dots + \mu_N$ .

(2) The Variance of the distribution for  $x$  is the sum of the Variances:  $V = V_1 + V_2 + V_3 + \dots + V_N$ .

(3) The distribution for  $x$  becomes Gaussian for large  $N$

# CLT - why?

Proof is fun but a bit long – see book

Illustration: Take uniform distribution in range 0 to 1..

One on its own is rectangular

Two added give a triangular distribution in range 0 to 2

Three added start with a concave parabola, switch to a convex one, flat at the top then back down

Twelve give something so Gaussian it's used to generate random numbers

(1) Is obvious.

(2) is simple and explains 'adding errors in quadrature.'

(1) and (2) do not depend on the form of the distribution

(3) Explains why Gaussians are 'normal'

If you find a distribution which is not Gaussian, there must be a reason  
Probably one contribution dominates

# Application of CLT

If a variable has a non-Gaussian pdf you can still apply parts (1) and (2): adding variances, using combination of errors, etc.

The only thing you can't do is equate deviations with confidence regions (68% within one sigma etc)

However your variable is probably intermediate and will be a contribution to some final result – Gaussian by (3). So carry on

*Non-Gaussian distributions hold no terrors*

# Errors from Likelihood

Estimate a model parameter  $M$  by maximising the likelihood

In the large  $N$  limit

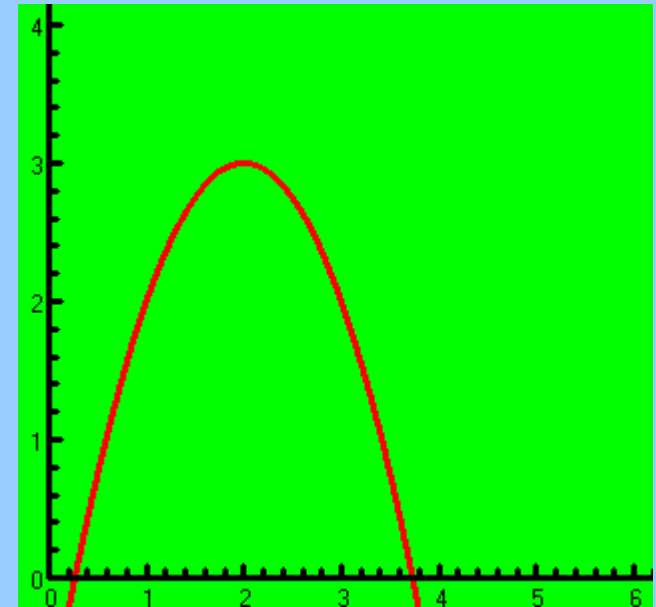
i) This is unbiased and efficient

ii) The error is given by

$$\frac{1}{\sigma^2} = -\left\langle \frac{d^2 \ln L}{dM^2} \right\rangle$$

iii)  $\ln L$  is a parabola

$$L = L_{max} - \frac{1}{2} C (M - \hat{M})^2$$



iv) We can approximate

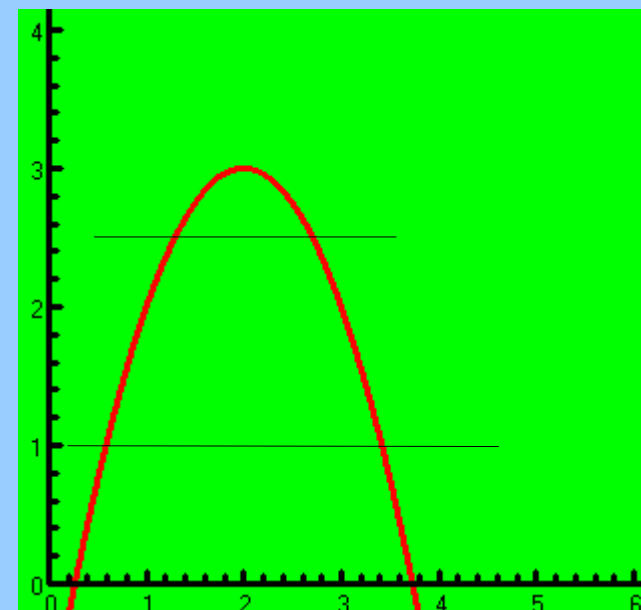
$$C \equiv \left. \frac{-d^2 \ln L}{dM^2} \right|_{M=\hat{M}} = -\left\langle \frac{d^2 \ln L}{dM^2} \right\rangle$$

v) Read off  $\sigma$  from  $\Delta \ln L = -1/2$

Take  $\Delta \ln L = -\frac{1}{2}$  for 68%  
CL ( $1\sigma$ )

$\Delta \ln L = -2$  for 95.4% CL ( $2\sigma$ )

Or whatever you choose  
2-sided or 1-sided





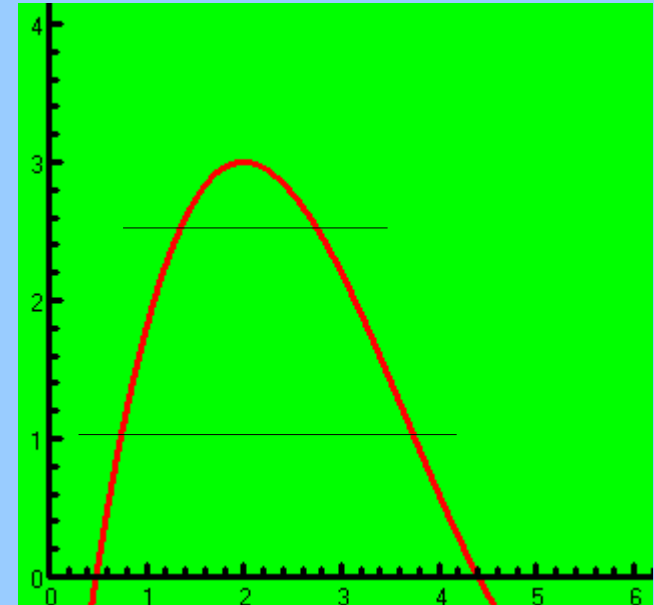
# For finite N

None of the above are true

Never mind! We could transform from  $M \rightarrow M'$  where it was parabolic, find the limits, and transform back

Would give  $\Delta \ln L = -1/2$  for 68% CL etc as before

Hence asymmetric errors



*Everybody does this*

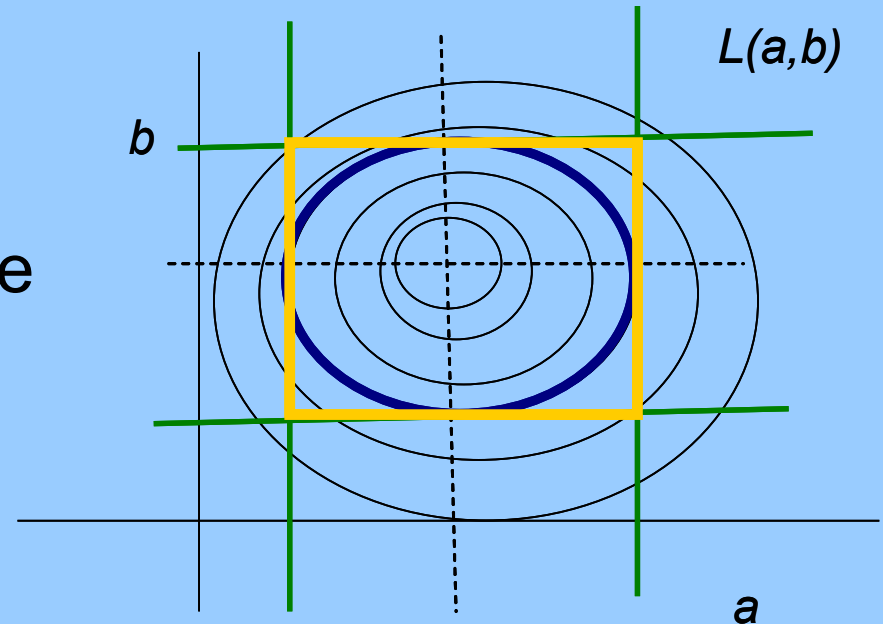
# Is it valid?

Try and see with toy model  
 (lifetime measurement)  
 where we can do the  
 Neyman construction  
 For various numbers of  
 measurements,  $N$ ,  
 normalised to unit lifetime  
 There are some quite severe  
 differences!

N	Exact		$\Delta \ln L = -\frac{1}{2}$	
	$\sigma_-$	$\sigma_+$	$\sigma_-$	$\sigma_+$
1	0.457	4.787	0.576	2.314
2	0.394	1.824	0.469	1.228
3	0.353	1.194	0.410	0.894
4	0.324	0.918	0.370	0.725
5	0.302	0.760	0.340	0.621
6	0.284	0.657	0.318	0.550
7	0.270	0.584	0.299	0.497
8	0.257	0.529	0.284	0.456
9	0.247	0.486	0.271	0.423
10	0.237	0.451	0.260	0.396
15	0.203	0.343	0.219	0.310
20	0.182	0.285	0.194	0.261
25	0.166	0.248	0.176	0.230

# More dimensions

Suppose 2 uncorrelated parameters,  $a$  and  $b$   
 For fixed  $b$ ,  $\Delta \ln L = -1/2$  will give 68% CL region for  $a$   
 And likewise, fixing  $a$ , for  $b$   
 Confidence level for square is  $0.68^2 = 46\%$   
 Confidence level for ellipse (contour) is 39%



Jointly,  $\Delta \ln L = -1/2$  gives 39% CL region  
 for 68% need  $\Delta \ln L = -1.15$

# More dimensions, other limits

Useful to write

$$-2\Delta\ln L = \chi^2$$

Careful! Given a multidimensional Gaussian,  $n$   
 $\ln L = -\chi^2/2$ . Hence can also use  $\Delta\chi^2=1$  for errors

But  $-2\Delta\ln L$  obeys a  $\chi^2$  distribution only in the large  $N$  limit...

Level is given by finding  $\chi^2$  such that  $P(\chi^2, N) = 1 - \text{CL}$

Generalisation to correlated gaussians is straightforward

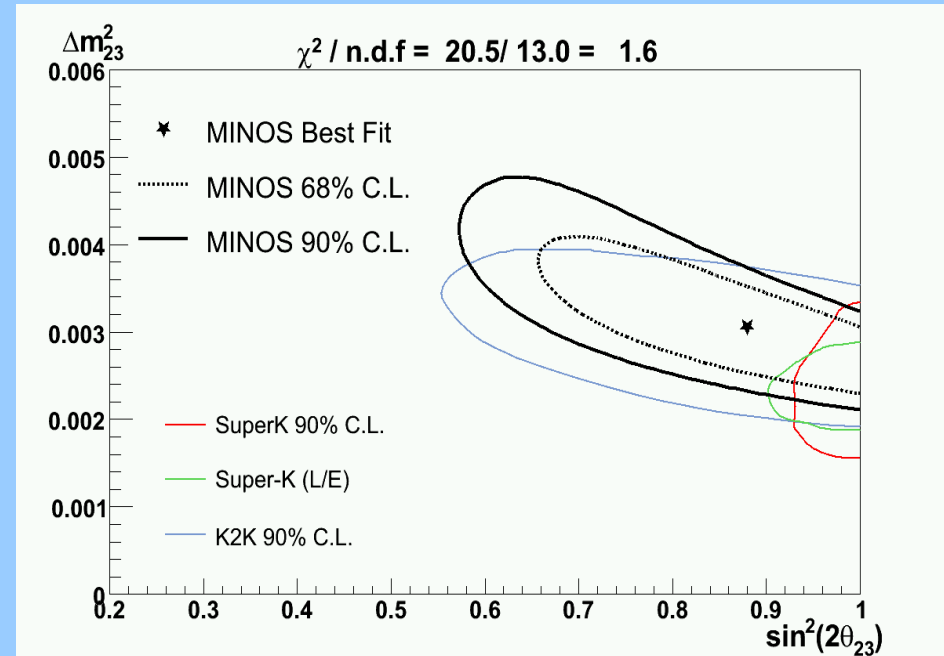
Generalisation to more variables is straightforward. Need the larger  $\Delta\ln L$

	68%	95%	99%
1	0.5	1.92	3.32
2	1.15	3.00	4.60
3	1.77	3.91	5.65

etc

# Small N non-Gaussian measurements

No longer  
 ellipses/ellipsoids  
 Use  $\Delta \ln L$  to define  
 confidence regions,  
 mapping out contours  
 Probably not totally  
 accurate, but universal





Given some  $f(x,y)$

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + 2\rho \left(\frac{\partial f}{\partial x}\right) \left(\frac{\partial f}{\partial y}\right) \sigma_x \sigma_y$$

$$\rho = \frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}}$$

Also matrices

$$V' = \tilde{G} V G$$

# Example

Collect  $N_T$  events

$N_F$  forwards

$N_B$  backwards

Want error on

$$R = N_F / N_T$$

Everything Poisson

F and B uncorrelated

F and T correlated

$$\text{Cov} = \langle N_F N_T \rangle - \langle N_F \rangle \langle N_T \rangle$$

$$= V(N_F) = N_F$$

using

$$\langle N_F N_B \rangle = \langle N_F \rangle \langle N_B \rangle$$

$$\rho = \sqrt{N_F / N_T}$$



Using  $R = N_F / N_T$

$$\begin{aligned} \sigma_R^2 &= \left(\frac{1}{N_T}\right)^2 N_F + \left(\frac{-N_F}{N_T^2}\right)^2 N_T + 2\sqrt{N_F/N_T} \left(\frac{1}{N_T}\right) \left(\frac{-N_F}{N_T^2}\right) \sqrt{N_F N_T} \\ &= \frac{N_F N_T + N_F^2 - 2 N_F^2}{N_T^3} = \frac{R(1-R)}{N_T} \end{aligned}$$

Using  $R = N_F / (N_F + N_B)$

$$\sigma_R^2 = \left(\frac{1}{N_T} - \frac{N_F}{N_T^2}\right)^2 N_F + \left(\frac{-N_F}{N_T^2}\right)^2 N_B = \left(\frac{N_B}{N_T^2}\right)^2 N_F + \left(\frac{N_F}{N_T^2}\right)^2 N_B = \frac{N_F N_B}{N_T^3} = \frac{R(1-R)}{N_T}$$

# Systematic Errors

“Systematic errors arise from neglected effects such as incorrectly calibrated equipment.”

Agree or disagree?

# Systematic Errors

“Systematic errors arise from neglected effects such as incorrectly calibrated equipment.”

**FALSE!**

A neglected effect is a **MISTAKE**

A **MISTAKE** is not an **ERROR**

(as we tell the undergraduates on day1)

So what are they?

# Systematic Errors

Analysis of your results involves a whole set of numerical factors: efficiencies, magnetic fields, dimensions, calibrations...

Occasionally these are implicit: these are especially dangerous

All these numbers have an associated uncertainty.

These uncertainties are the systematic errors. They obey all the usual error laws, but they affect all measurement

The magnetic field in  $p = 0.3 \text{ B R}$

Calorimeter energy calibration

'Jet energy scale'

Detector efficiency

...

If you can't think of them for your experiment, ask a colleague with a talent for destructive criticism. (There are plenty around)

Effect of uncertainty in B on the error matrix for two momentum measurements

$$V = \begin{pmatrix} 0.3^2 B^2 \sigma_1^2 + 0.3^2 R_1^2 \sigma_B^2 & 0.3^2 R_1 R_2 \sigma_B^2 \\ 0.3^2 R_1 R_2 \sigma_B^2 & 0.3^2 B^2 \sigma_2^2 + 0.3^2 R_2^2 \sigma_B^2 \end{pmatrix}$$

Errors on  $p_1$  and  $p_2$  as given by simple combination of errors.

Also covariance /correlation term. Errors in B effect both momentum measurements the same way

## More complicated...

Many properties of the reconstruction don't work through simple algebra.

Example: background to your signal simulated by Monte Carlo containing several (?) adjustable parameters...

Work numerically. Run standard MC, then adjust parameter by  $+\sigma$  and repeat,  $-\sigma$  and repeat. Read off error from shift in result.

If you can convince yourself that the 3 points are a straight line then do so.

## ***Don't sweat the small stuff!***

Errors add in quadrature. Go for the biggest. Reducing small errors still further is a waste of your energy:

$$\sqrt{(10^2+2^2)}=10.20$$

$$\sqrt{(10^2+1^2)}=10.05$$



Check your result by altering features which should make no (significant) difference. This adds to its credibility

Run on subsets of the data (time etc)

Change cuts on quality and kinematic quantities

Check that a full blown analysis on simulated data returns the physics you put in

Repeat until you (and you supervisor and review committee) really believe

# 'Significant'

Repeating with some difference in technique will give a different result.

You have to decide whether this is significant.

“Within Errors” may be overgenerous as results share the same data (or some of it)

Subtraction in quadrature is one way:

Basic result  $12.3 \pm 0.4$ . Check  $11.7 \pm 0.5$

Compare difference 0.6 against  $\sqrt{.5^2 - .4^2} = .3$

If the analysis passes the check with a small difference

***Tick the box and move on***

Do not fold that small difference into the systematic error

If the analysis fails the check

- 1) Check the check
- 2) Check the analysis and find the problem
- 3) Maybe convince yourself that this 'harmless' change could cause a systematic shift and devise an appropriate error

Do not fold the difference into the systematic error

# Source of confusion

## Two tables – similar yet different

Vary

- Energy scale
- Mag field
- Trigger effcy
- MC parameters
- ...

and include  
results in  
systematic  
errors

Vary by  
1 sigma

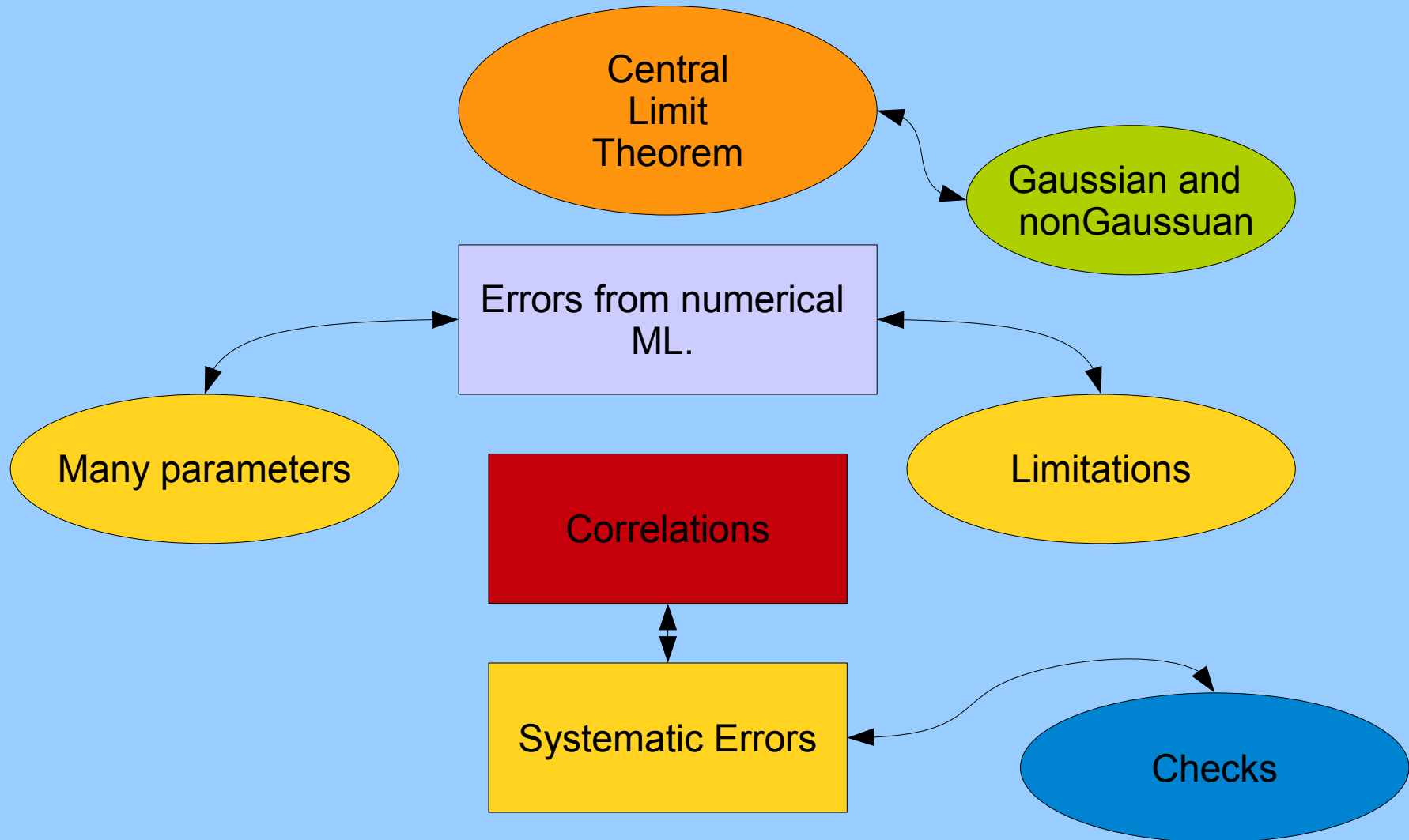
Vary

- Energy cut
- Lepton quality
- isolation
- ...

But do not  
include results  
in systematic  
errors

Vary by  
Arbitrary  
amount

# Summary



Statistics is a science, not an art. There is a reason for everything. Understand what you are doing and why.

Cheap computing is opening many new ways of doing things. Use it!

There is a lot of bad practice out there. Do not take the advice of your supervisor/senior colleague/professor as infallible.