Introduction to Statistics for High Energy Physics

Stéphane Monteil, Clermont University,

LPC-IN2P3/CNRS/UCA

Programme co-funded by the European Union

Disclaimer for this lecture.

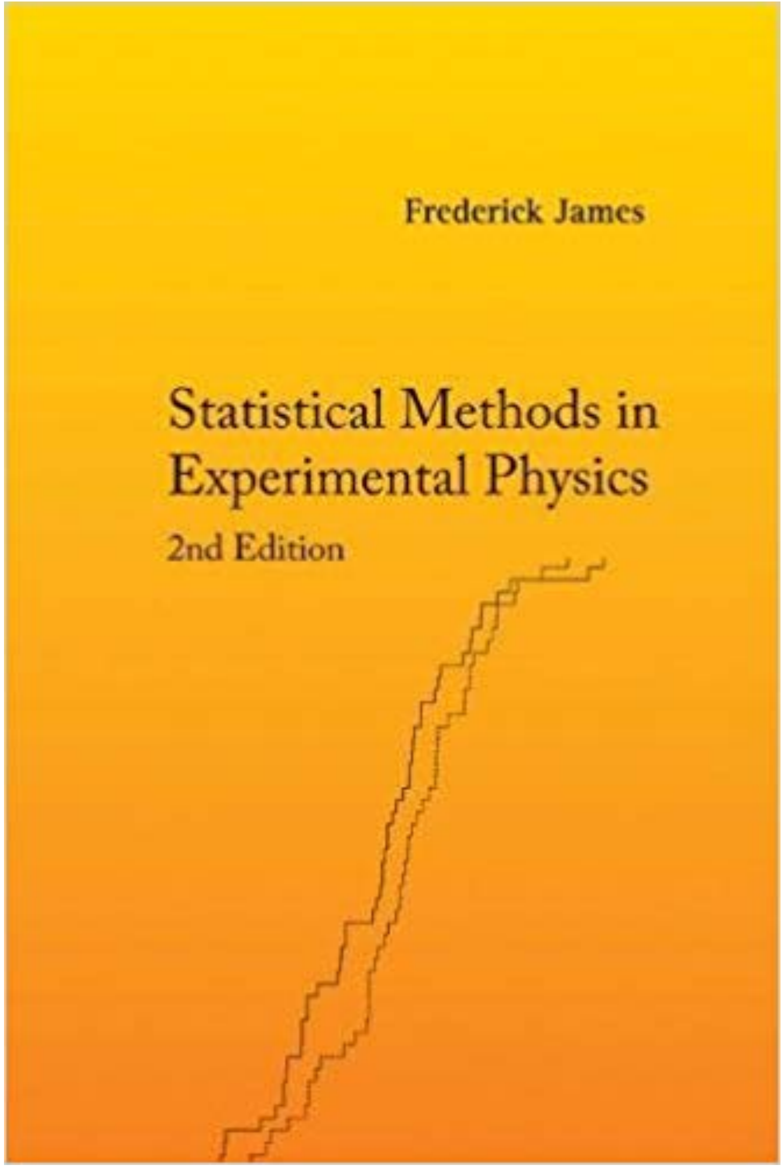
- I'm serving here as a substitute. Please consult the reference statistics lecture in TESHEP by **Jonas Rademacker.** This introduction is too rapid.

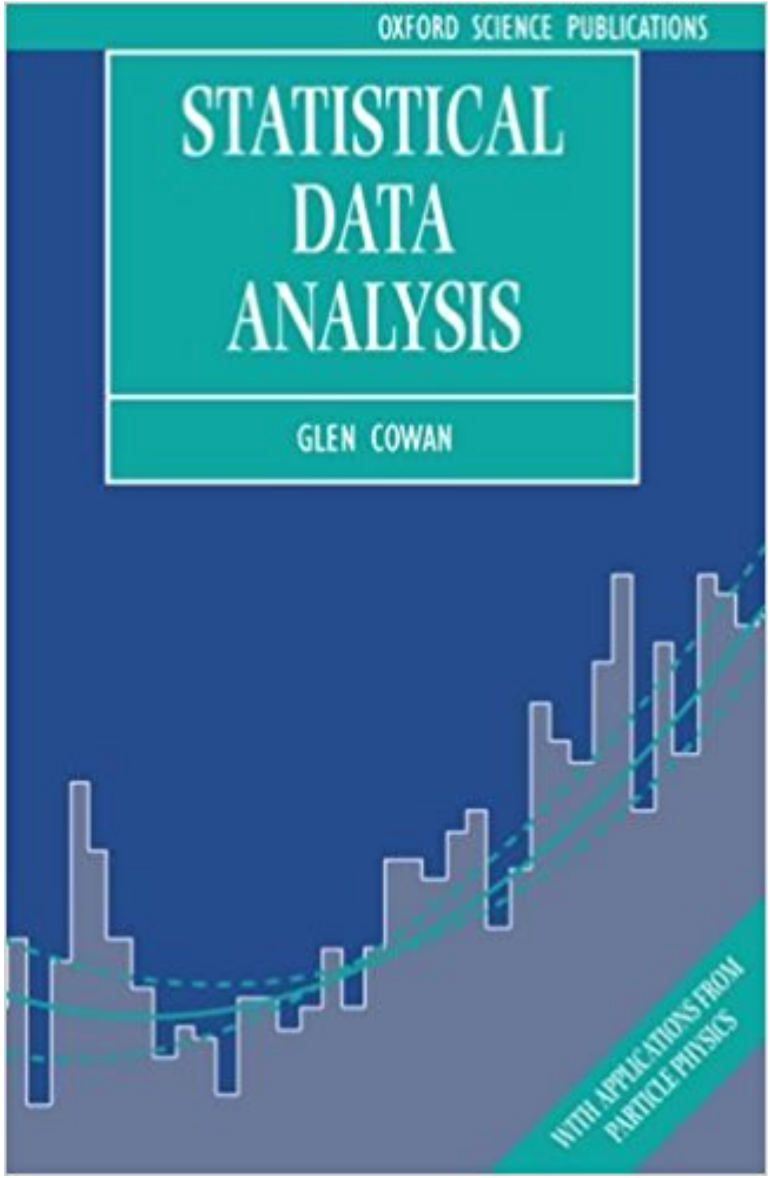
- Materials borrowed from Clermont's colleagues (E. Busato, O. Deschamps, R. Madar, L. Serlet). © Statistics@Clermont

- Get to HEP-oriented fundamental books (everything useful is in there):

  - Frederick James
  - Roger Barlow
  - Glen Cowan

Bibliography advises

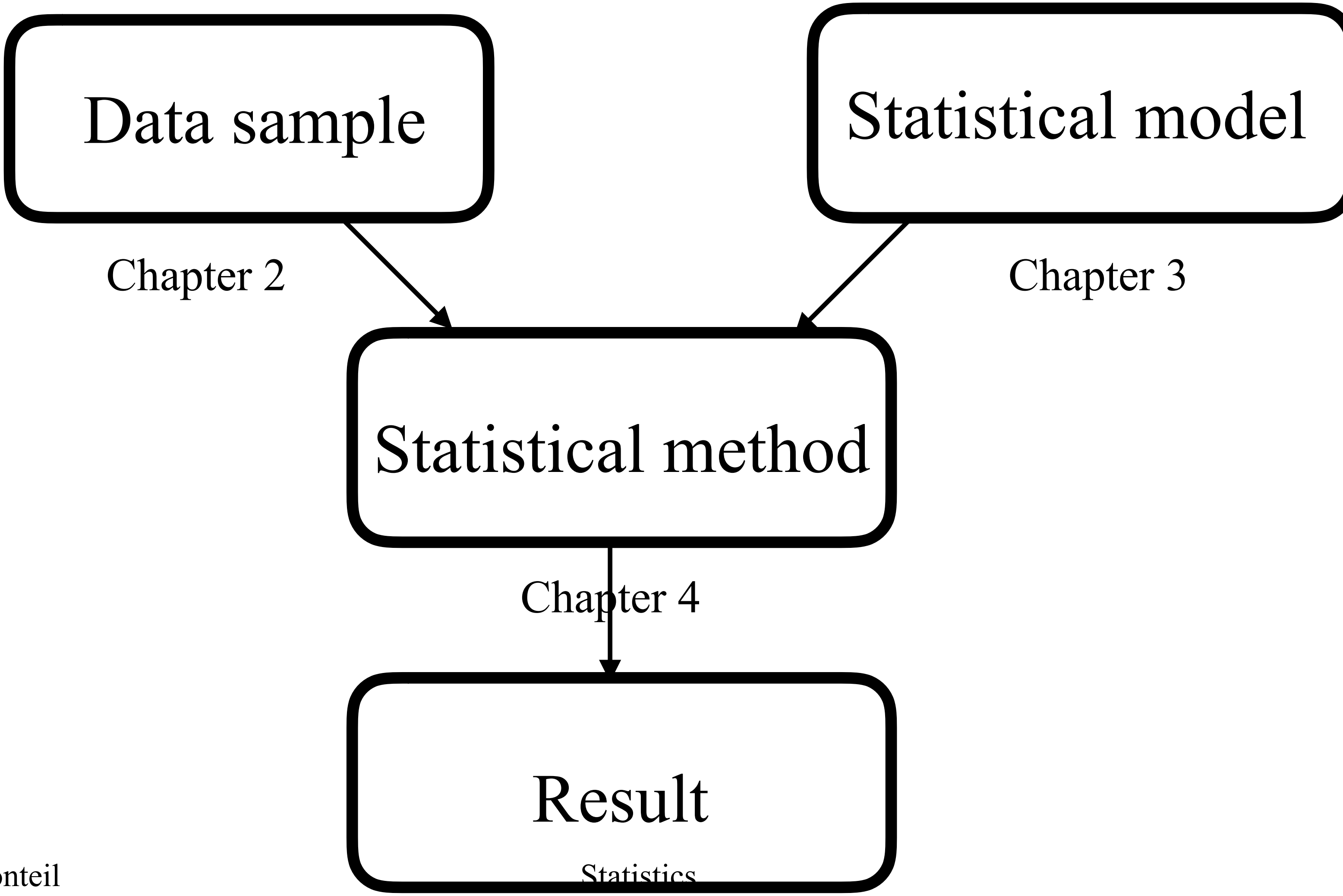




Lecture @ TESHEP2023

Outline of the lecture

a brief motivation and then



0) The shortest History of Probability and Statistics ever

- The calculus of probabilities might be born in the 17th century in the european *salons* where gambling games were played and then studied.

- One of the first reference (french-centered) is the address of *Chevalier de Méré* [1610-1685] to *Blaise Pascal* (1623-1662): is it more probable to obtain (at least) a 6 in four dice throws or a double 6 in 24 two-dices throws? [Get your exercise done and sit on the shoulders of the giants]

- The 18th century provided the limit theorems with noticeably *Bernouilli* and *De Moivre.*

- The 19th century witnessed further progresses and the establishment of the fundamental probability laws we‘re using on an everyday basis *(Gauss, Laplace, Poisson …)*

0) The shortest History of Probability and Statistics ever

- But the modern theory is born in the 20th century thanks to *Kolmogorov* and the foundations of the theory of the measurement by *Borel* and *Lebesgues.*

- The theory of randomness is continuously developing since then as an intense field of research, which applications are everywhere in our everyday life. This lecture might well be written by a conversational assistant …

- It happens it is not (it would be much better if it were).

0) and motivations: why do you care about randomness?

- A la Cyrano de Bergerac
- Greedy: because gambling, sportive bets and games of chances are making people rich, though not those who are playing …
- Cautious: because we want to evaluate risks: climate, weather, seismic activities.
- Cautious and greedy: covering the risks is a natural human characteristics. Those actually covering the risk, insurance companies and finance investors are using financial mathematics to maximise their profits.
- Curious: the fundamental properties of Nature are definite numbers (think of the mass or the charge of the electron). Measurements of nature are on the contrary coming with biases, estimated with uncertainties. This is why mastering randomness is important!

0) and motivations: why do you care about randomness?

- Final disclaimer of this introduction:
- We are usually taught in HEP-oriented lectures that statistics is somehow an art; there are several methods at hand; nothing is forbidden if you state what you've done; etc…
- All that is probably true … but one should not forget the following
- Statistics is a branch of mathematics:
  - It is axiomatic !
  - Most of the methods are proven !
  - Asymptotic limits are known !
- In the following all approximations are mine.

Chapter I: Elements of probability theory

Chapter I: Elements of probability theory

- Let’s make some warm-up probability gymnastics: and come back to Blaise Pascal (I’m contractually obliged to cite him repeatedly)

- Let’s denote \Omega the universe of the possibles containing all possible finite outcomes \omega_i

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$$

- In the finite probabilistic model, each sub-set of \Omega will be called an event (and can be written literally or mathematically)

- The sum of all probabilities \omega_i must obey: $$\sum_{i=1}^{N} p(\omega_i) = 1$$

- Let’s make some warm-up probability gymnastics: and come back to Blaise Pascal (I’m contractually obliged to cite him repeatedly)

- Then if we select a sub-set of the outcomes:

$$\forall A \in \mathcal{P}(\Omega),\ \mathbb{P}(A) = \sum_{\omega_i \in A} p(\omega_i)$$

- To calculate any probability, the knowledge of $p(w_i)$ is required.In the simple case of an identical probability for each outcome, you need to know how to count !

- The probability of a set of occurrences is the number of those occurrences realised divided by all the possibilities.

Chapter I: Elements of probability theory

- Probability modelling in simple cases: a random experiment in which there are a finite number of outcomes.

- Few counting highlights:

  - n persons in this lecture room, k chairs, *permutation* of k among n

$$A_n^k = \frac{n!}{(n-k)!}$$

  - If you don't care who is seated where, *combination*:

$$C_n^k = \binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{A_n^k}{k!}.$$

- Back to Blaise Pascal (I'm contractually obliged to cite him repeatedly)

- Is it more probable to obtain (at least) a 6 in four dice throws or a double 6 in 24 two-dices throws ?

  - First case:
    - the universe is the cartesian product $\Omega = \{1, 2, 3, 4, 5, 6\}^4$
    - the event "at least one six" has the complementary event "no 6", hence:

$$\Omega = \{1, 2, 3, 4, 5, \cancel{6}\}^4$$

$$P(\text{"at least a 6"}) = 1 - P(\text{"no 6"}) = 1 - \frac{5^4}{6^4}.$$

Chapter I: Elements of probability theory

- Back to Blaise Pascal (I'm contractually obliged to cite him repeatedly)
- Is it more probable to obtain (at least) a 6 in four dice throws or a double 6 in 24 two-dices throws ?
  - Second case:
    - the universe is $\Omega = (\{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\})^{24}$
    - the event "at least (6,6)" has the complementary event "no (6,6)", which is the 24-uplets not having (6,6) hence:

$$P(\text{"at least a } (6,6\text{"}) = 1 - P(\text{"no } (6,6\text{"}) = 1 - \frac{35^{24}}{36^{24}}.$$

Chapter I: Elements of probability theory

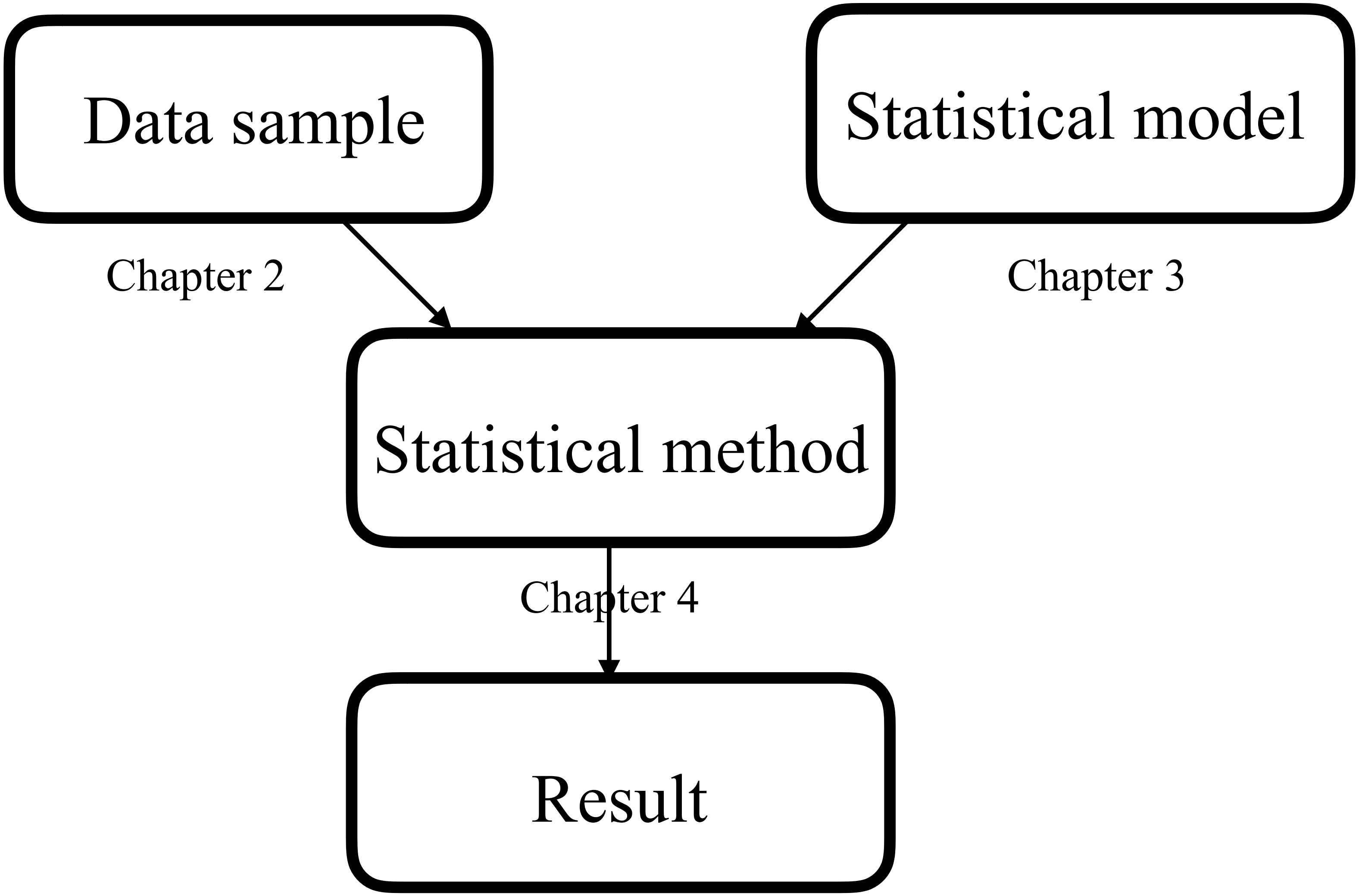
- Exercise: what is the probability that the event “A: at least two of us in this room do share the same anniversary date” is realised?
- Guess?
- It is the complement of the event “ Abar: noone in the room are sharing the same anniversary date”
- The latter event is the permutation of the n persons among the m days of the year (say 365 to make it simple):
- The universe of possibilities has a cardinal: m^n.

Chapter I: Elements of probability theory

- Exercise: what is the probability that the event “A: at least two of us in this room do share the same anniversary date” is realised?
- Guess?
- It is the complement of the event “ Abar: noone in the room are sharing the same anniversary date”
- The latter event is the permutation of the n persons among the m days of the year (say 365 to make it simple):
- The universe of possibilities has a cardinal: m^n.

If we are 40 in the room, the probability is ~90%!

Outline of the lecture



Chapter II: Random var. as elements of the data sample

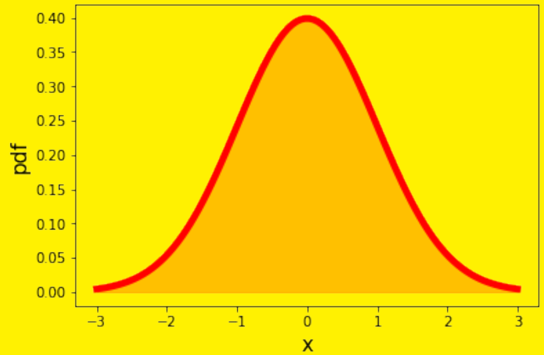
- A sample is chosen to represent the population one wants to study
- A sample is a set of measured random variables. But what is a random (aleatory) variable ?
- It is a quantity which is not certain (owns an intrinsic randomness)
- Could be misleading since it is neither random nor variable !
- It is rather a function from possible outcomes in a sample space to a measurable space.
- e.g.
  - the result of heads or tails of several coin flips
  - the value of an observable (to which an uncertainty is attached)

Chapter II: Random var. as elements of the data sample

- The random variables can be either discrete or with density

Continuous random variable

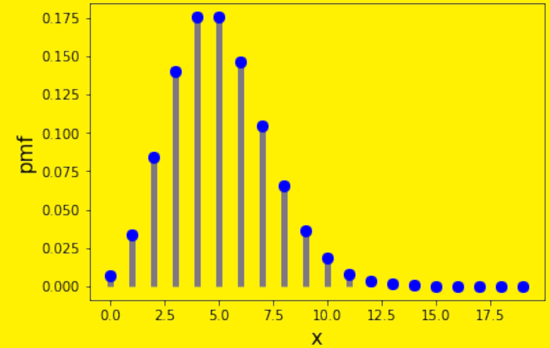
- Characterized by pdf: $f_X(x;\theta)$



- $P(x \leq X \leq x + dx) = f_X(x;\theta)\mathrm{d}x$
- $\mathbb{E}[X] = \int x f_X(x;\theta)\mathrm{d}x$

Discrete random variable

- Characterized by pmf: $p_X(x;\theta)$



- $p_X(x;\theta) =$ prob. that $X = x$
- $\mathbb{E}[X] = \sum_i x_i p_X(x_i;\theta)$

Other important relations:

$\rightarrow$ $\mathrm{var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$

$\rightarrow$ cdf: $F(t;\theta) = P(X \leq t;\theta)$

Chapter II: Random var. as elements of the data sample

- The properties of the sample (representing the population)

  - **Sample mean**: $M = \frac{1}{n}\sum_{i=1}^{n} X_i$

  - **Sample median**: assuming $X_1 < X_2 < \ldots < X_n$
    - median$= \frac{X_{n/2} + X_{1+n/2}}{2}$ (n even)
    - median$= X_{(n+1)/2}$ (n odd)

  - **Sample variance**:

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - M)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - M^2 \quad \text{“biased”}$$

or

$$S_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M)^2 \quad \text{“unbiased”}$$

  - **Sample covariance matrix**: $Q = \frac{1}{n-1}\sum_{i=1}^{n}\left(\vec{X}_i - \vec{M}\right)\left(\vec{X}_i - \vec{M}\right)^T$

- If you know the probability law that governs the random variables of interest, you know the exact properties of the population

Population parameters (not subject to fluctuations) : expected value, (co)variance, moments, …

- ***Expected value*** : $\mathbb{E}[X] = \mu$

Linearity
$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$
$$\mathbb{E}[aX] = a\,\mathbb{E}[X]$$
$$\mathbb{E}[X+a] = \mathbb{E}[X] + a$$
$Non-multiplicativity\ (a\ priori) : \mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$

- **Variance** : $var[X] = \mathbb{V}[X] = \sigma^2 = \mathbb{E}[(X-\mu)^2]$

$Koenig-Huygens\ formula : \mathbb{V}[X] = \sigma^2 = \mathbb{E}[X^2] - \mu^2$

- **Covariance** : $cov(X,Y) = \sigma_{XY} = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$

$\mathbb{V}[X]$ = cov(X,X)
If cov(X,Y)=0 then $\mathbb{V}[X+Y] = \mathbb{V}[X] + \mathbb{V}[Y]$ (Bienaymé form.)

- $(\boldsymbol{order\ n})\ \boldsymbol{moments}$: $\mathbb{E}[X^n] = m_n$

$m_0 = 1$
$m_1 = \mathbb{E}[X] = \mu$
$m_2 = \mathbb{E}[X^2]$
$\mathbb{V}[X] = \sigma^2 = m_2 - \mu^2$

- ***central moments*** : $\mathbb{E}[(X-\mu)^n] = \mu_n$

$\mu_0 = 1\ ;\ \mu_1 = 0$
$\mathbb{V}[X] = \mu_2 = \sigma^2$

- ***standardized moments*** : $\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^n\right] = \tilde{\mu}_n$

$\tilde{\mu}_0 = 1\ ;\ \tilde{\mu}_1 = 0\ ;\ \tilde{\mu}_2 = 1$
$\tilde{\mu}_3 = \gamma_1$ (skewness) ; $\tilde{\mu}_4 = \beta_2 = \gamma_2 + 3$ (Kurtosis)

- A word about independence

Independence

- By definition, X and Y are independent if:

$$\mathbf{f_{XY}(x,y) = f_X(x) f_Y(y)}$$

Or equivalently: $f_X(x|y) = f_X(x)$ and $f_Y(y|x) = f_Y(y)$

- **Remark:** Dependence $\neq$ Correlation
  - Correlation measured by covariance:

$$\text{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

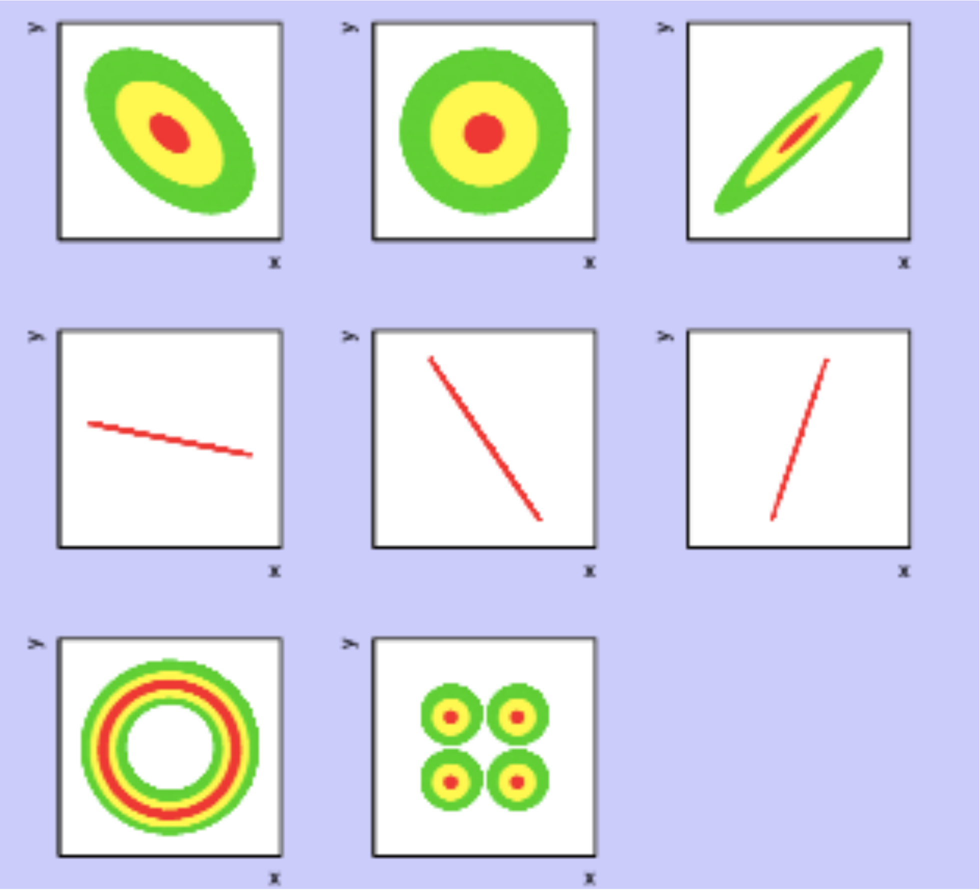
  - Uncorrelated when $\text{cov}(X,Y) = 0$

- **Independence** $\Rightarrow$ **Uncorrelation** (but not the contrary)

- A useful estimator is the linear(!) correlation $\rho_{xy} = \frac{\text{Cov}(x,y)}{\sigma_x \sigma_y}$

Chapter II: Random var. as elements of the data sample

- Exercise: how much correlated?



Chapter II: Random var. as elements of the data sample

- The standard comment in all statistics lectures: Correlation is not Causation! And statistical properties would never capture the causality.

Chapter II: some useful p.m.f.

- Some examples of canonical probability law or density: Binomial.
- Two (binomial) or many (multinomial) discrete outcomes (success/failure); (-1,1).

Binomial and multinomial distributions

☐ **Binomial**: $P(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}$

Properties:

→ $\mathbb{E}[k] = np$

→ $\mathrm{var}[k] = np(1-p)$

☐ **Multinomial**: $P(n_1,\cdots,n_m;n,p_1,\cdots,p_m) = \frac{n!}{n_1!\cdots n_m!} p_1^{n_1}\cdots p_m^{n_m}$

where

- m: number of possible results in a trial
- n_i: number of results of type i ($i \in [1;m]$), $\sum n_i = n$
- p_i: probability that result in a trial is of type i

Properties:

→ $\mathbb{E}[n_1] = np_i$

→ $\mathrm{var}[n_i] = np_i(1-p_i)$

→ $\mathrm{cov}(n_i,n_j) = -np_ip_j$

Chapter II: some useful p.m.f.

- Binomial: Head / Tail; Success / Failure; Triggered event / rejected event

The adequate probability law to deal with the uncertainties of an efficiency determination. Check the variance.

☐ **Binomial**: $P(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}$

- Multinomial: not two possibilities but a finite number > 2. e.g. you make a selection of a decay mode and you classify them in intervals of their invariant mass: that's an histogram !

☐ **Multinomial**: $P(n_1,\cdots,n_m;n,p_1,\cdots,p_m) = \frac{n!}{n_1!\cdots n_m!} p_1^{n_1}\cdots p_m^{n_m}$

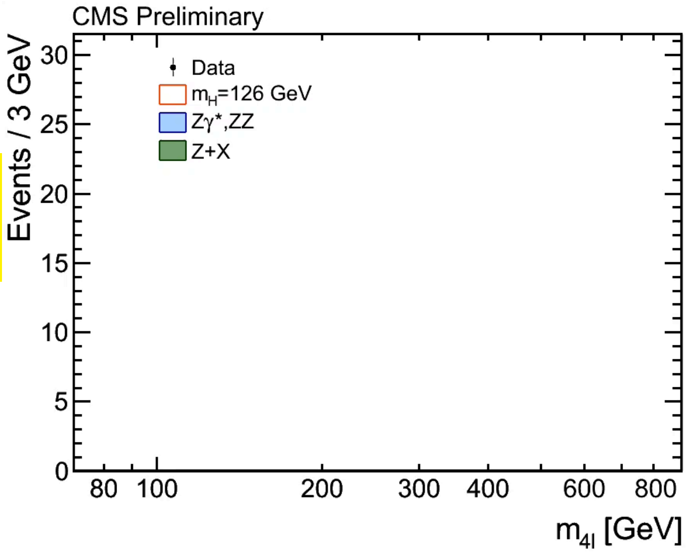
Chapter II: some useful p.m.f.

- Binomial: Head / Tail; Success / Failure; Triggered event / rejected event

The adequate probability law to deal with the uncertainties of an efficiency determination. Check the variance.

☐ **Binomial**: $P(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}$

- Multinomial: not two possibilities but a finite number > 2. e.g. you make a selection of a decay mode and you classify them in intervals of their invariant mass: that's an histogram !



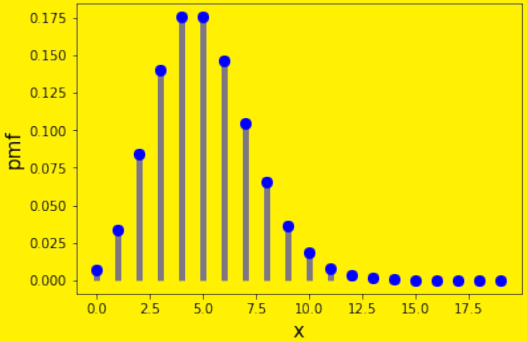
☐ **Multinomial**: $P(n_1,\cdots,n_m;n,p_1,\cdots,p_m) = \frac{n!}{n_1!\cdots n_m!} p_1^{n_1}\cdots p_m^{n_m}$

Chapter II: some useful p.m.f.

- Some examples of canonical probability law or density: Poisson
- Deals w/ rare numbers.

Poisson distribution

$$P(n;\nu) = \frac{\nu^n}{n!} e^{-\nu}$$



Properties:

- $\mathbb{E}[n] = \nu$
- $\mathrm{var}[n] = \nu$

- ☐ Poisson = limit of binomial when $n \to \infty$ and $p \to 0$

$$\binom{n}{k} p^k (1-p)^{n-k} \xrightarrow[\substack{n\to\infty \\ p\to 0}]{} \frac{\nu^n}{n!} e^{-\nu} \quad \text{with} \quad \nu = np$$

- ☐ Poisson aka the "law of rare events"

Chapter II: some useful p.m.f.

- Some examples of canonical probability law or density: Poisson

- Exercise:
  - the SM predicts that in your experiment (that you designed carefully such that there is no background), you shall see 5 neutrinos.

  - You observe 2.

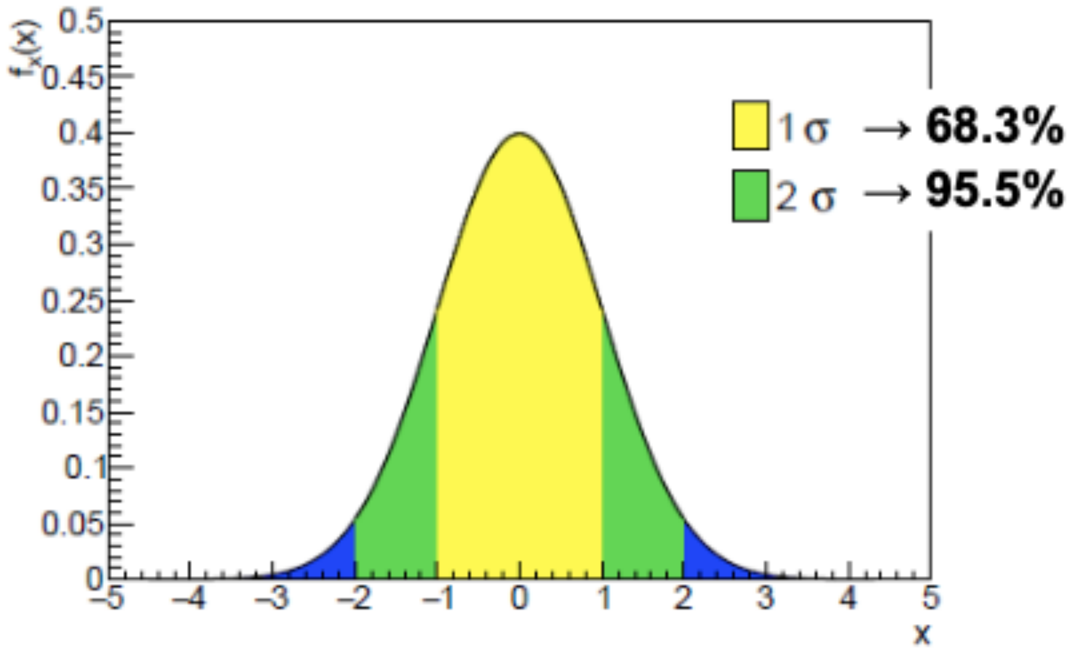
  - How likely is it?
  - Poisson law of parameter 5

$$P(X = k, \nu) = \frac{k^{\nu} e^{-\nu}}{k!}$$

$$P(X = 2) = \frac{2^5 e^{-5}}{2!} \sim 10.8\%$$

Chapter II: some useful p.d.f.

- The most important probability density function of all: Gaussian law



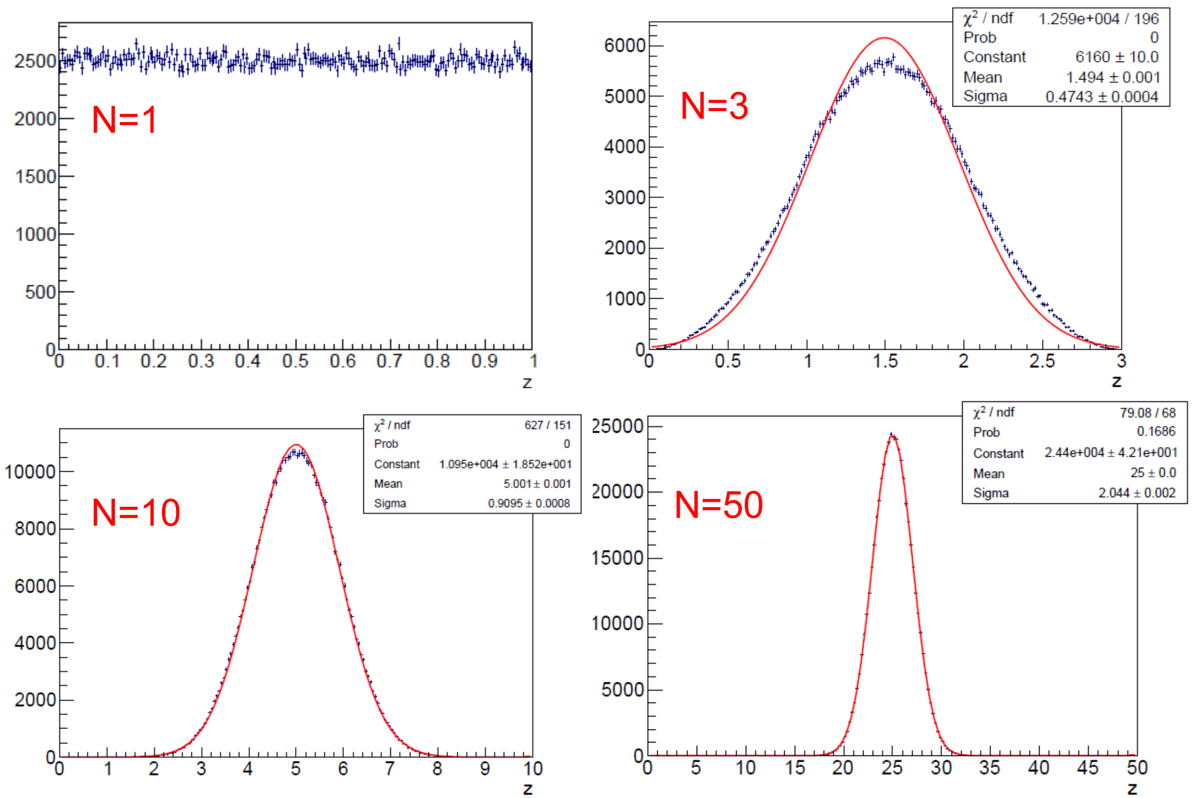
$$f(x;\mu;\sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2})$$

Mean

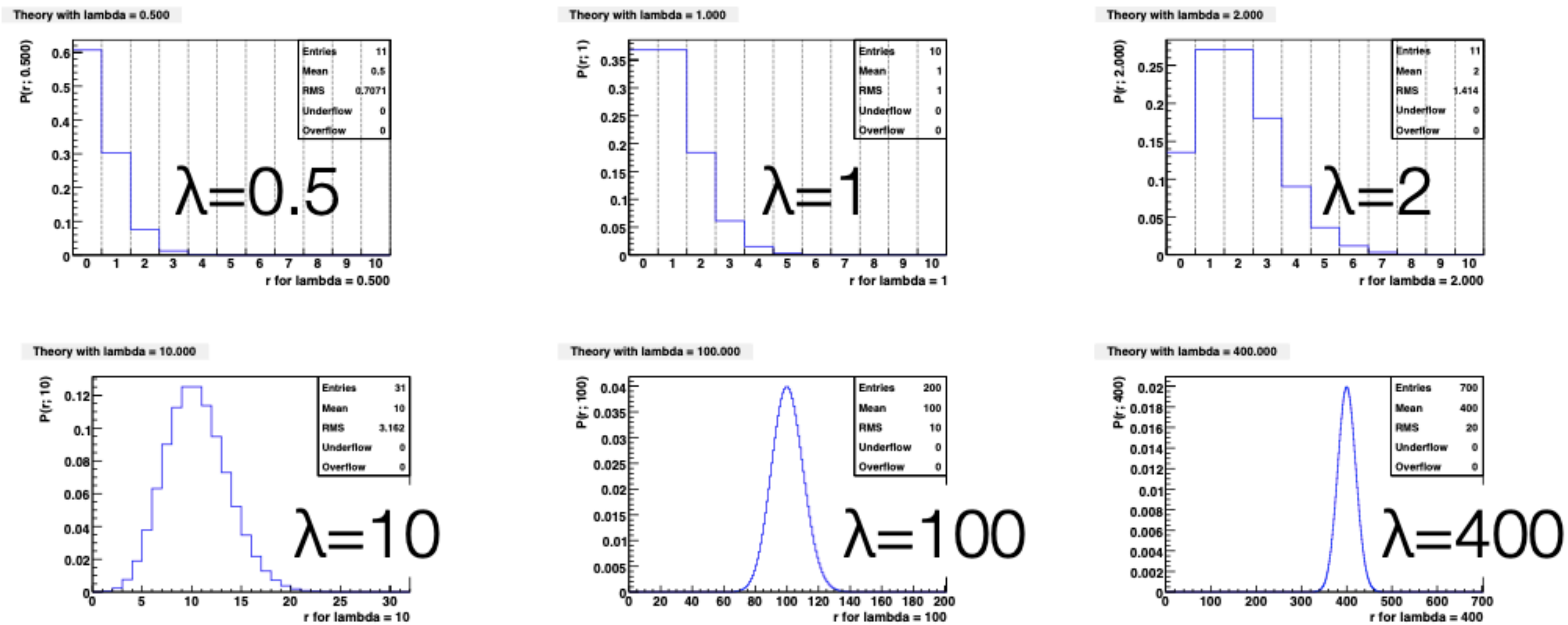
Variance

- Central limit theorem: pseudo-experiment proof.
- We consider a vector of variables x_i uniformly distributed in [0,1]
- We build the distribution of the N repetition of the x_i

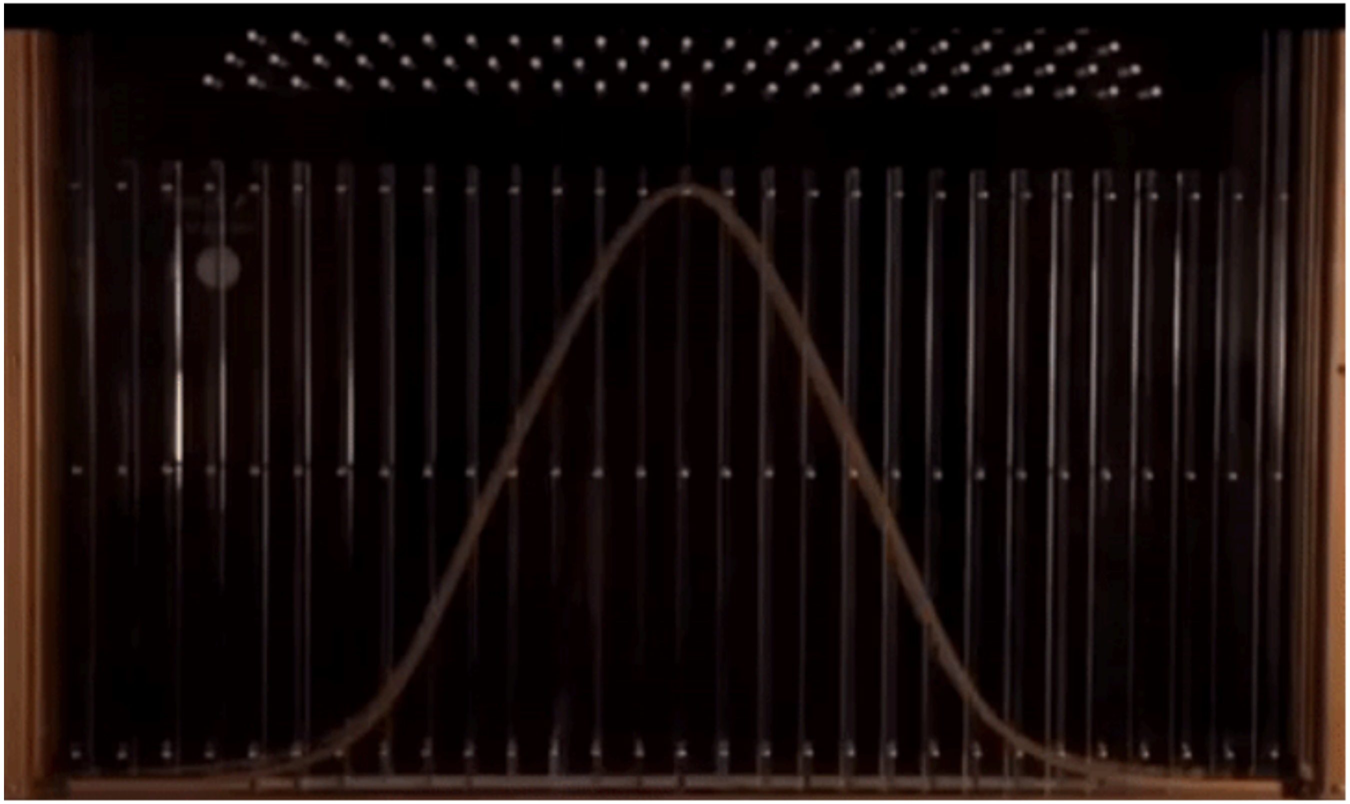
$$z = \sum_{i=1}^{i=N} x_i$$



- This is true for almost all distributions ! Here the Poisson's law captured from Jonas pseudo-exepriments.



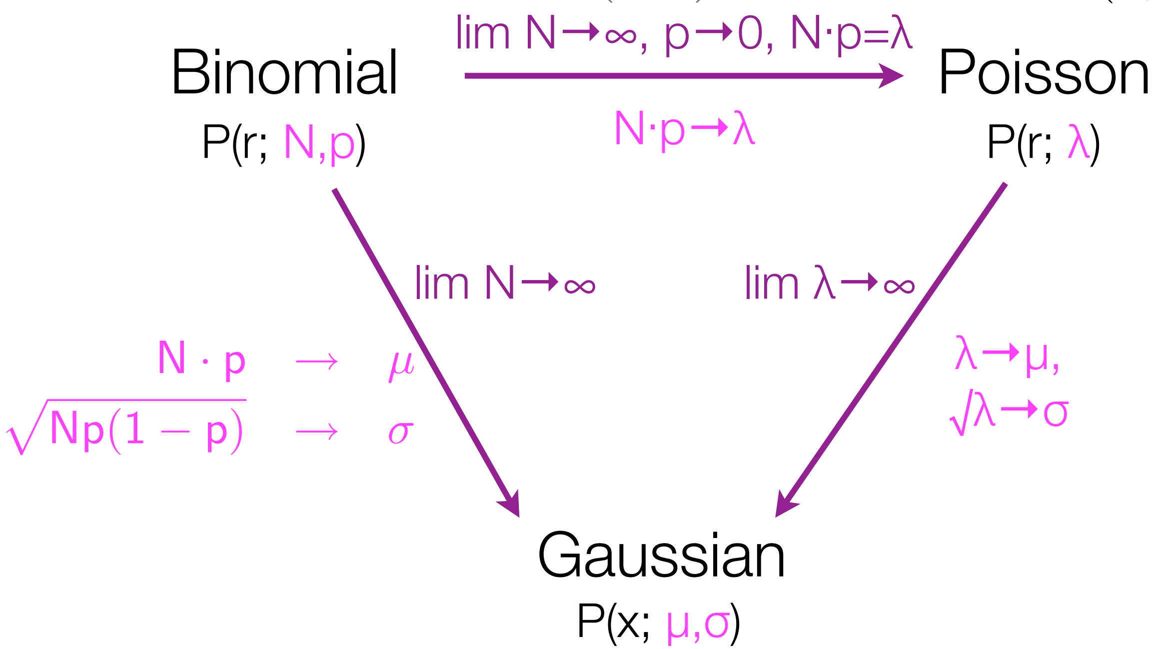
- Central limit theorem: if you prefer a hardware proof:
- https://en.wikipedia.org/wiki/Galton_board



Trinity

$$P(r; N, p) = p^r (1-p)^{N-r} \binom{N}{r}$$

$$P(r;\lambda) = e^{-\lambda}\frac{\lambda^r}{r!}$$



$$g(x;\mu,\sigma = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Chapter II: Covariance and error matrices

- What happens if we are dealing with multidimensional samples, e.g. with *N* measurements of *M* variables? The (likely) gaussian blur of the measurements becomes encoded into covariance (error) matrices, the diagonal elements of them dealing with the actual variances of the observables and the off-diagonal terms.

Covariance matrix (aka error matrix) of sample $\{\vec{x_i}\}, \mathrm{i} = 1..\mathrm{N}$

- Real, symmetric, N×N matrix of the form:

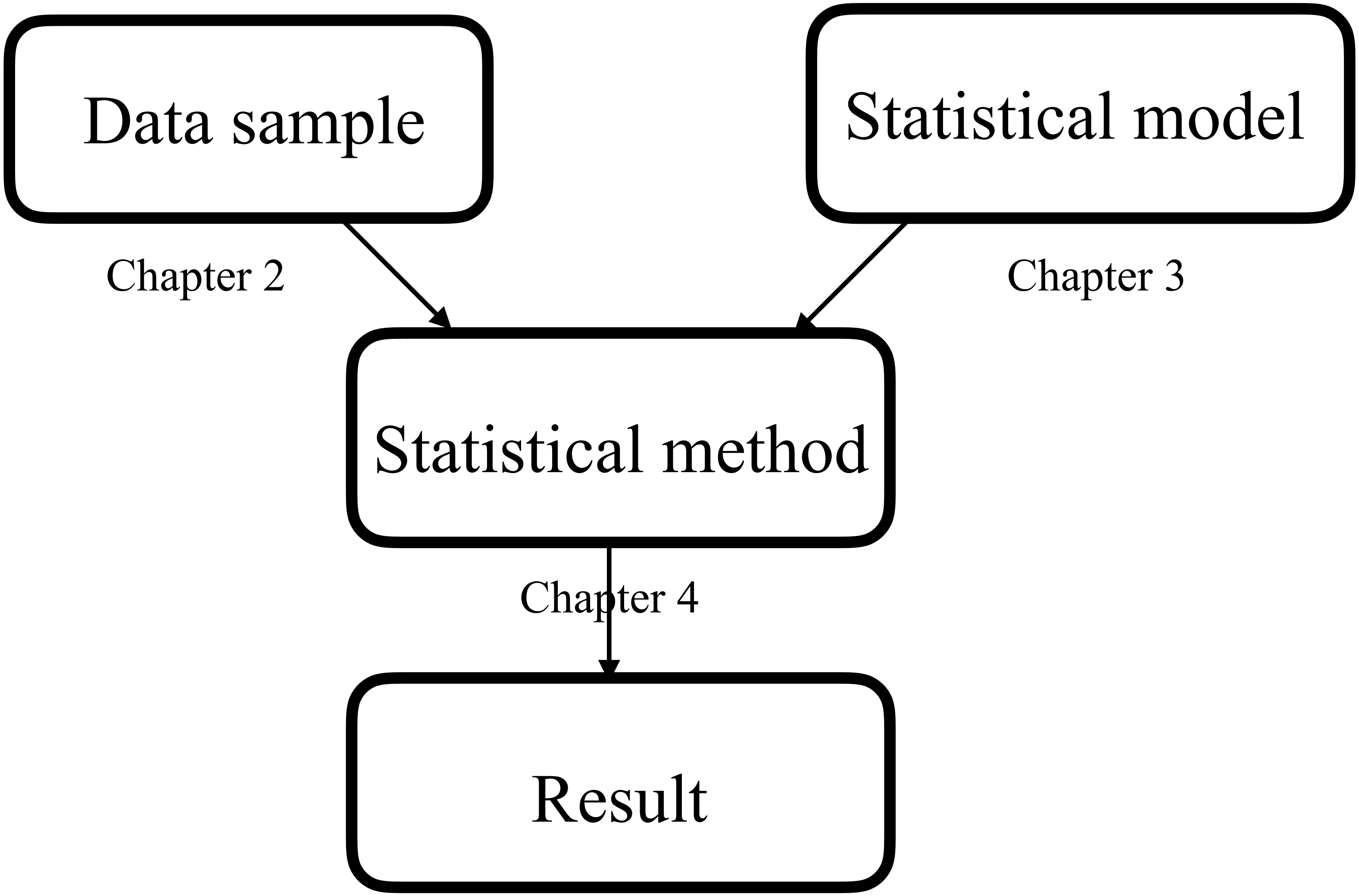
$$C = \begin{pmatrix} \mathrm{cov}(x_1, x_1) & \cdots & \mathrm{cov}(x_1, x_N) \\ \vdots & \mathrm{cov}(x_i, x_j) & \vdots \\ \mathrm{cov}(x_N, x_1) & \dots & \mathrm{cov}(x_N, x_N) \end{pmatrix} = \begin{pmatrix} {\sigma_1}^2 & \cdots & \rho_{1N}\sigma_1\sigma_N \\ \vdots & \rho_{ij}\sigma_i\sigma_j & \vdots \\ \rho_{N1}\sigma_N\sigma_1 & \dots & {\sigma_N}^2 \end{pmatrix}$$

Correlation matrix: $\rho = \begin{pmatrix} 1 & \cdots & \rho_{1N} \\ \vdots & 1 & \vdots \\ \rho_{N1} & \dots & 1 \end{pmatrix}$

Example of usage of covariance matrix:

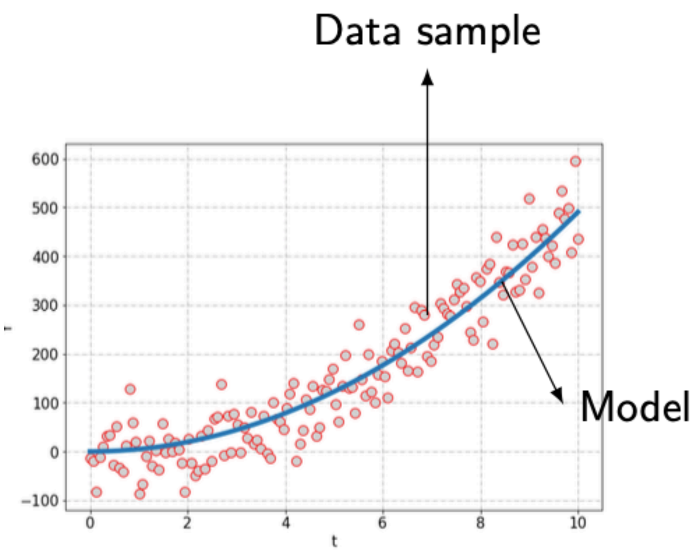
- Transformation of input variables
- Error propagation
- Combination of correlated measurements

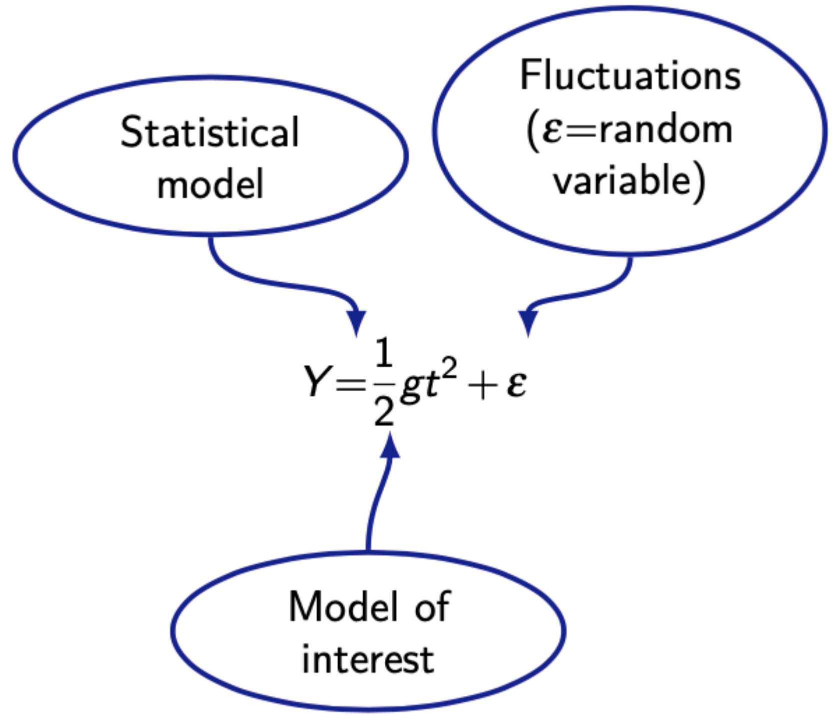
Chapter III: Statistical model



Chapter III: Statistical model

- What are we speaking of?





Stat. model = model of interest + fluctuation model (describes fluctuations inherent to measurement)

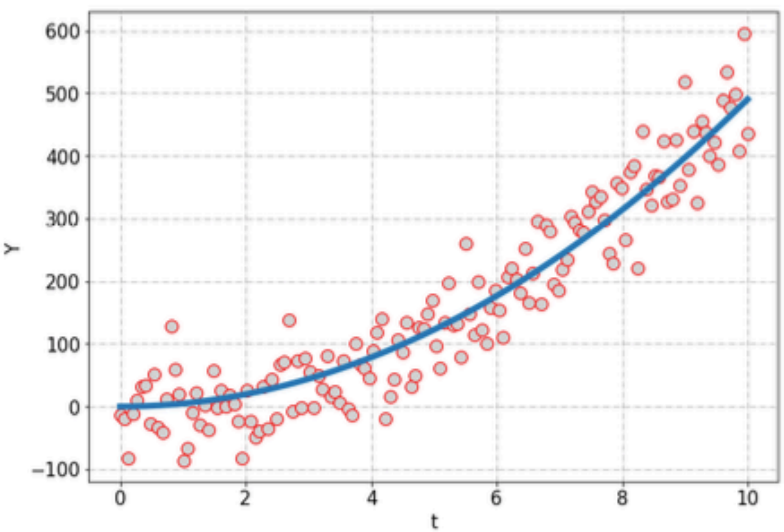
What is the measurement performed?

- **Ingredients and vocabulary.**

☐ Statistical model describing free fall measurement:

$$Y = \frac{1}{2} g t^2 + \varepsilon$$

- Y: observable (dependent variable)
- g: parameter of interest
- t: explanatory variable
- ε: can depend on some nuisance parameters



Chapter III: Statistical model: the likelihood

- Let's write initial **maths**

☐ Suppose you have a sample $X = (X_1, \cdots, X_n)$

☐ The stat. model can be written

$$f_{X_1 \cdots X_n}(x_1, \cdots, x_n; \theta) = f_{X_1}(x_1|x_2, \cdots, x_n; \theta) \times f_{X_2}(x_2|x_3, \cdots, x_n; \theta) \times \cdots \times f$$

☐ **For iid variables:**

$$f_{X_1 \cdots X_n}(x_1, \cdots, x_n; \theta) = f_X(x_1; \theta) \times f_X(x_2; \theta) \times \cdots \times f_X(x_n; \theta)$$

$$\Rightarrow \boxed{f_{X_1 \cdots X_n}(x_1, \cdots, x_n; \theta) = \prod_{i=1}^{n} f_X(x_i; \theta)}$$

☐ If the X_i's are iid normal variables:

$$\text{stat. model} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

$$= \frac{1}{(\sqrt{2\pi}\sigma)^n} exp\left(-\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

Oh! $-\chi^2/2$

- The statistical model is called the likelihood (identity). There are then several convenient ways to express it

  - $\ln \mathscr{L}$
  - $-\ln \mathscr{L}$
  - $\ln \mathscr{L}$ (or $-\ln \mathscr{L}$) with "constant terms" removed (constant terms = terms not depending on the parameters)

$\mathscr{L}(\theta; x)$

parameters (can have multiple-components)

observables (can have multiple-components)

- **Extending the likelihood:**

  - ☐ Size of the sample n often not constant but follows Poisson distribution: $n \sim \mathrm{Pois}(\nu)$

  - ☐ In such cases, the likelihood function has to be **extended**:

$$\mathscr{L}_{\mathrm{ext}}(\theta, \nu; x, n) = \frac{\nu^n}{n!} e^{-\nu} \times \mathscr{L}(\theta; x)$$

For iid case:

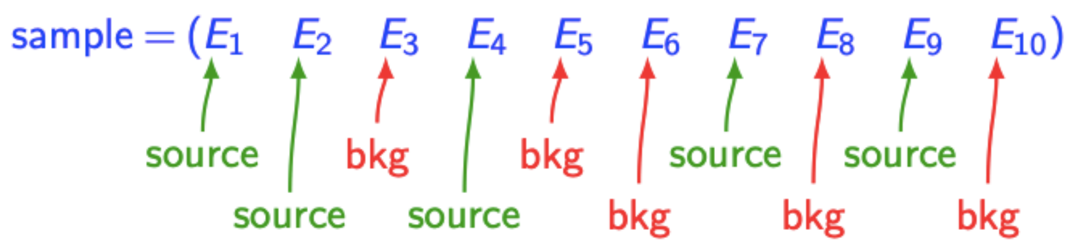
$$\mathscr{L}_{\mathrm{ext}}(\theta, \nu; x, n) = \frac{\nu^n}{n!} e^{-\nu} \times \prod_{i=1}^{n} \mathscr{L}(\theta; x_i)$$

extended likelihood — extended term — likelihood

⚠ Term likelihood may be used to denote extended likelihood
→ Should be clear from context whether we're talking about likelihood or extended likelihood

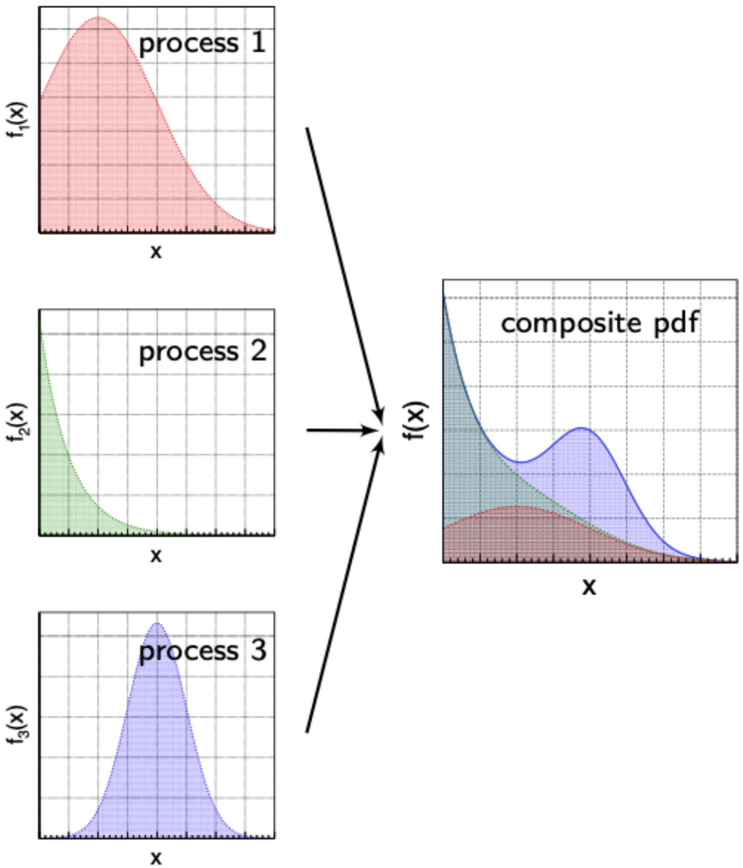
- Extending the likelihood to **multiple components:**

  - ☐ In realistic cases, samples are often **composite**
  - ☐ **Composite sample** = sample in which events can come from different origins
  - ☐ **Examples:**
    - Weights in a sample including men and women
    - Measurement of radioactive source:



  - ☐ Composite samples are said to be made of a **mixture of events**
  - ☐ Composite samples must be described by **composite stat. models** (or **composite likelihoods**)

- Extending the likelihood to **multiple components:**



- **Composite pdf**:

$$f(x;\{\mu_p\},\theta) = \frac{\sum_{p=1}^{P} \mu_p f_p(x;\theta)}{\sum_{p=1}^{P} \mu_p}$$

  where:
  - p: process index
  - μ_p: expected number of elements in sample from process p
  - $f_p(x;\theta)$: pdf for process p

- **Remark:** the μ_p's are often unknown

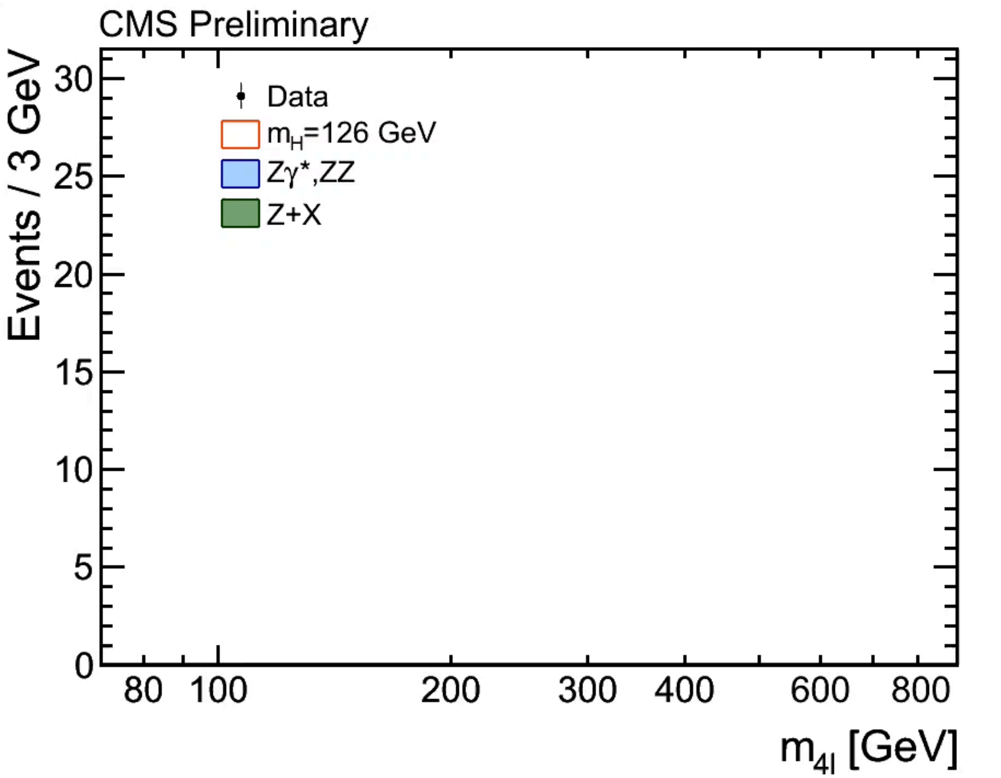
  $\rightarrow$ Determining them can be one of the objective

Chapter III: Statistical model: the likelihood extended

- Extending the likelihood to **multiple components:**

Chapter III: Statistical model: the likelihood extended

- Extending the likelihood to **multiple components:**



Chapter III: Statistical model: the likelihood extended

- And the parameter estimators (by a slight anticipation)
- The likelihood has deep, sound, extensive, **good** mathematical properties.
- The estimator maximising it owns those as well:

The ML method provides estimators with nice properties:

- **Functional invariance:**

$$\hat{\tau}_{\mathrm{ML}} = \tau(\hat{\theta}_{\mathrm{ML}})$$

- **ML estimators are consistent**
- If an unbiased and efficient estimator exists, then
  - → It is found by the ML method
  - → It is unique
  - → Its variance is

$$\mathrm{var}[\hat{\tau}_{\mathrm{ML}}] = -\frac{(\partial\tau/\partial\theta)^2}{\left.\frac{\partial^2 \ln \mathscr{L}}{\partial\theta^2}\right|_{\theta=\hat{\theta}_{\mathrm{ML}}}}$$

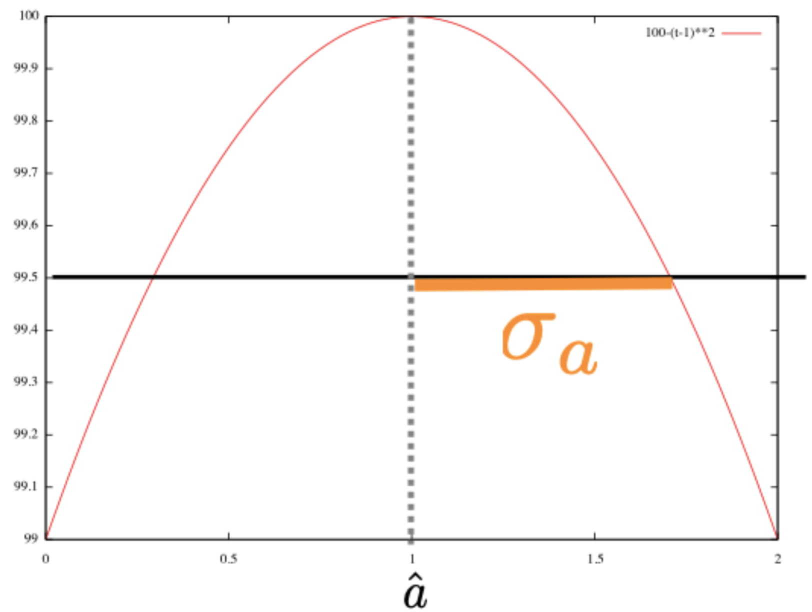
Rao-Cramer-Frechet Theorem (Special dedication to Pierre)

- ML estimators are **asymptotically normal**
- ML estimators are **asymptotically efficient**

- And the parameter estimators (by a slight anticipation) on the practical side: Maximise, scan the different values.
- Get the central value and the confidence intervals: Jonas illustration

$$\ln \mathcal{L} = -\frac{(a-\hat{a})^2}{2\sigma_a^2} + (\text{meaningless constant})$$

Asympotic limit
(Gaussian)



$$\Delta \ln \mathcal{L} = \frac{1}{2}$$

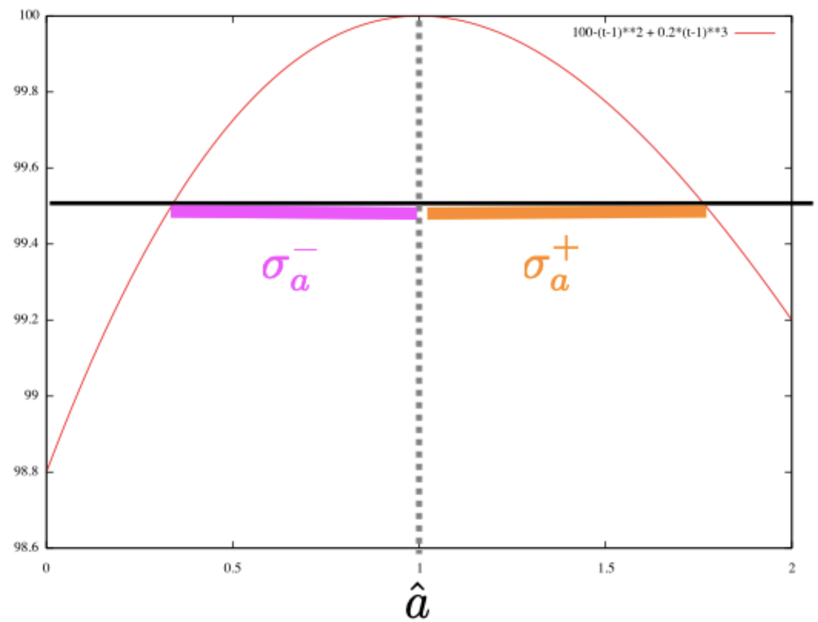
Chapter III: Statistical model: the likelihood extended

- And the parameter estimators (by a slight anticipation) on the practical side: Maximise, scan the different values.

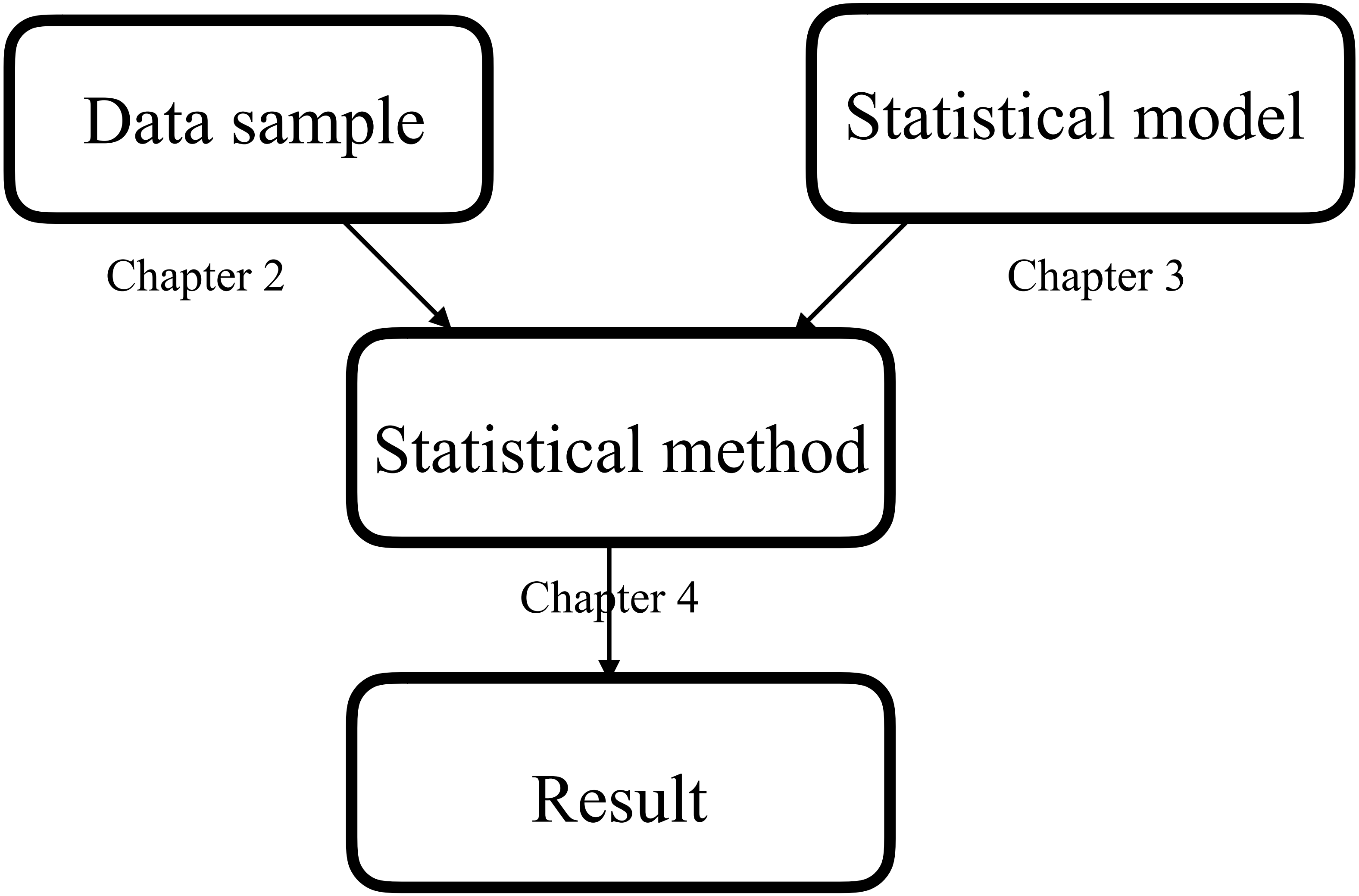
- Get the central value and the confidence intervals: Jonas illustration

$$a = \hat{a}^{+\sigma_a^+}_{-\sigma_a^-}$$

Asympotic limit
(e.g. Poisson)
Just quote asymmetric
uncertainties

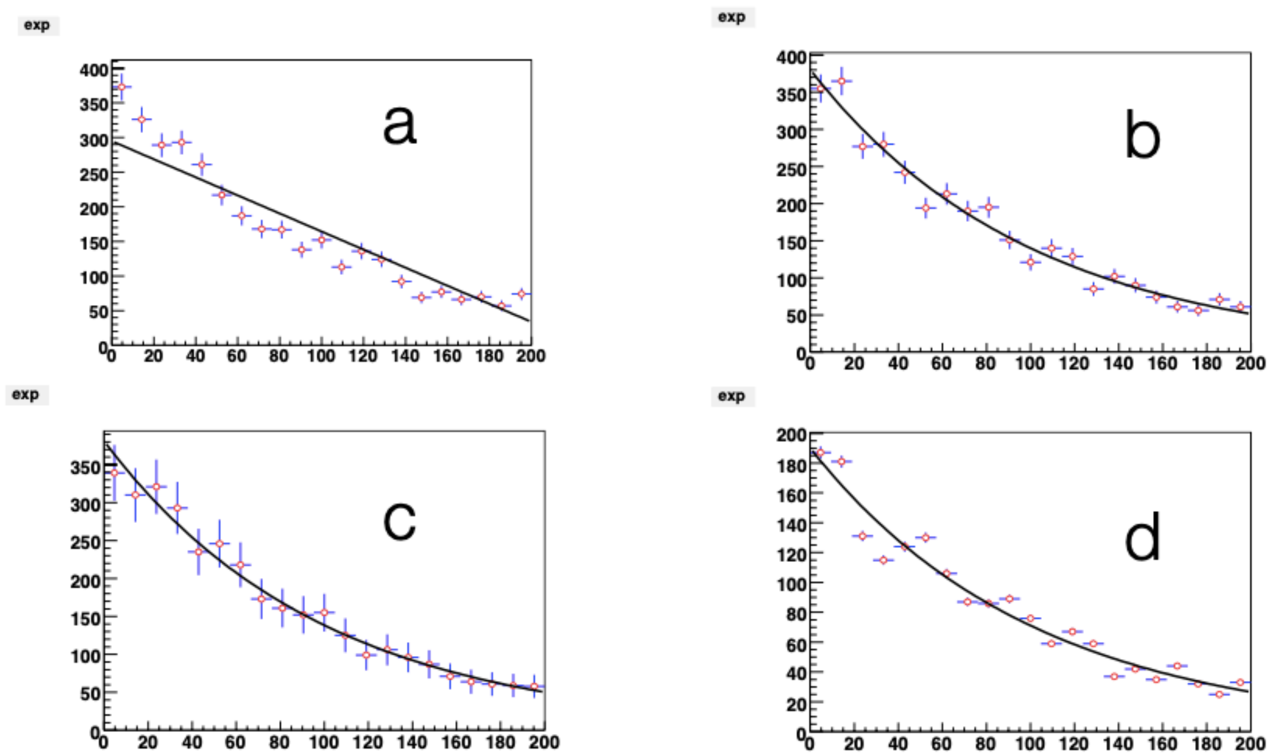


$$\Delta \ln \mathcal{L} = \frac{1}{2}$$


Data sample
Statistical model
Chapter 2
Chapter 3
Statistical method
Chapter 4
Result

- Is my model of interest (embodied in the statistical model through the likelihood) making sense? Let’s consider Jonas’s example with b-flavoured particles proper time fits. Exponential probability density function is your model of interest.

- **First, which fit is making sense?**



Chapter IV: Statistical method (Inferences)

- **It is therefore very useful to have an estimator of the quality of the fit.**

- No easy going implementation of a goodness of fit test with unbinned (event by event) likelihood fit. Use the associated binned histogram you use for visualisation. And compute the chi^2 !

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - f_i)^2}{\sigma_i^2}$$

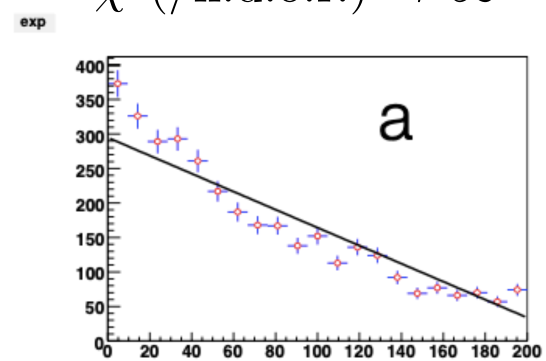
- What would be its value if the fit is correctly behaving?

- How to get a universal estimator (adapted to any distribution characteristics?) **Divide by the number of degrees of freedoms - 1.**

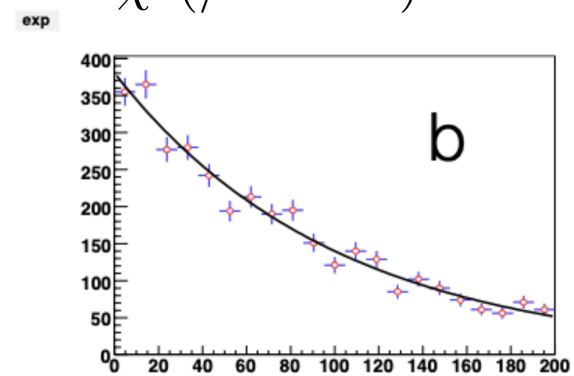
Chapter IV: Statistical method (Inferences)

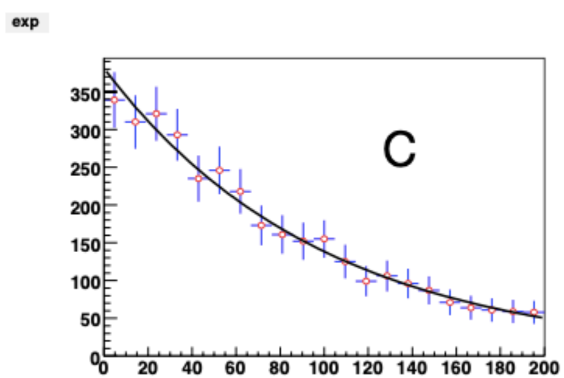
- **which fit is making sense?**

$$\chi^2(/\text{n.d.o.f.}) \to \infty$$

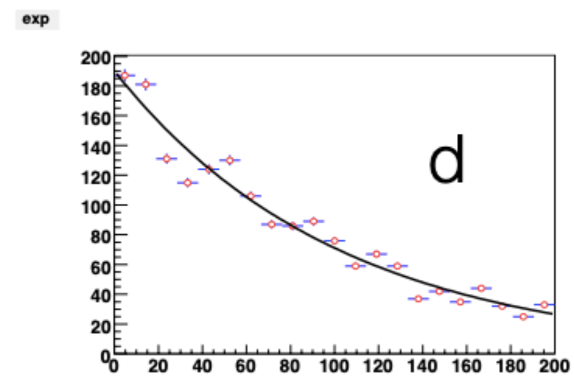


$$\chi^2(/\text{n.d.o.f.}) \sim 1$$





$$\chi^2(/\text{n.d.o.f.}) \sim 0$$

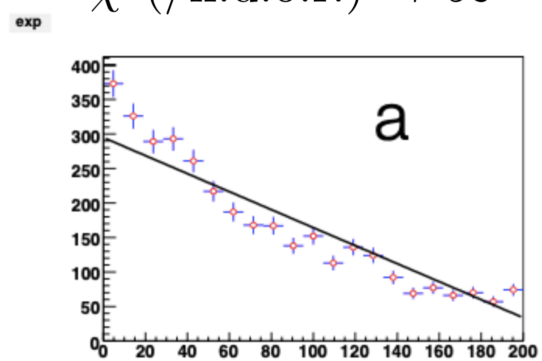


$$\chi^2(/\text{n.d.o.f.}) \to \infty$$

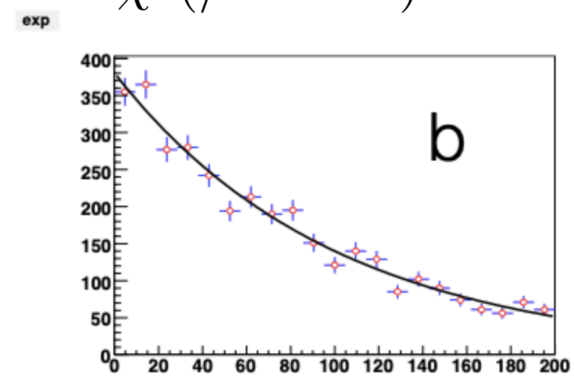
- Not a serious quantitative inference yet, just a check to proceed further!

- **which fit is making sense?**

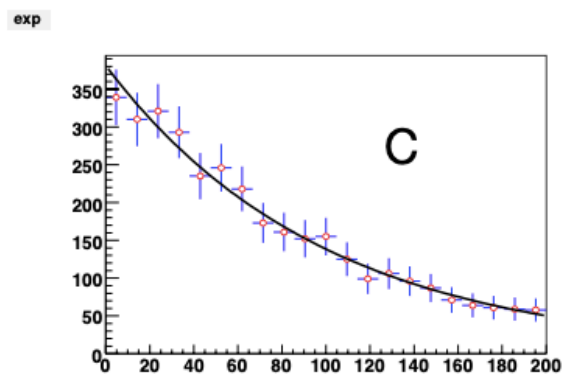
$\chi^2(/\text{n.d.o.f.}) \to \infty$



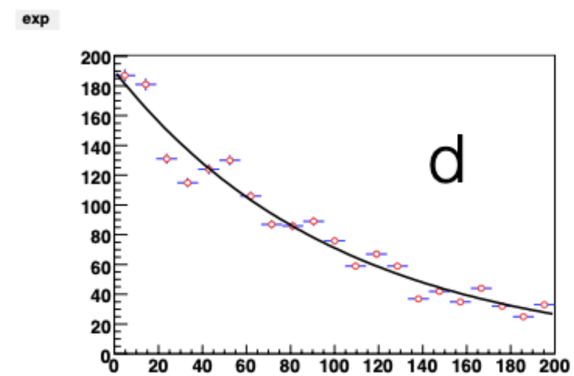
$\chi^2(/\text{n.d.o.f.}) \sim 1$



GOOD



$\chi^2(/\text{n.d.o.f.}) \sim 0$



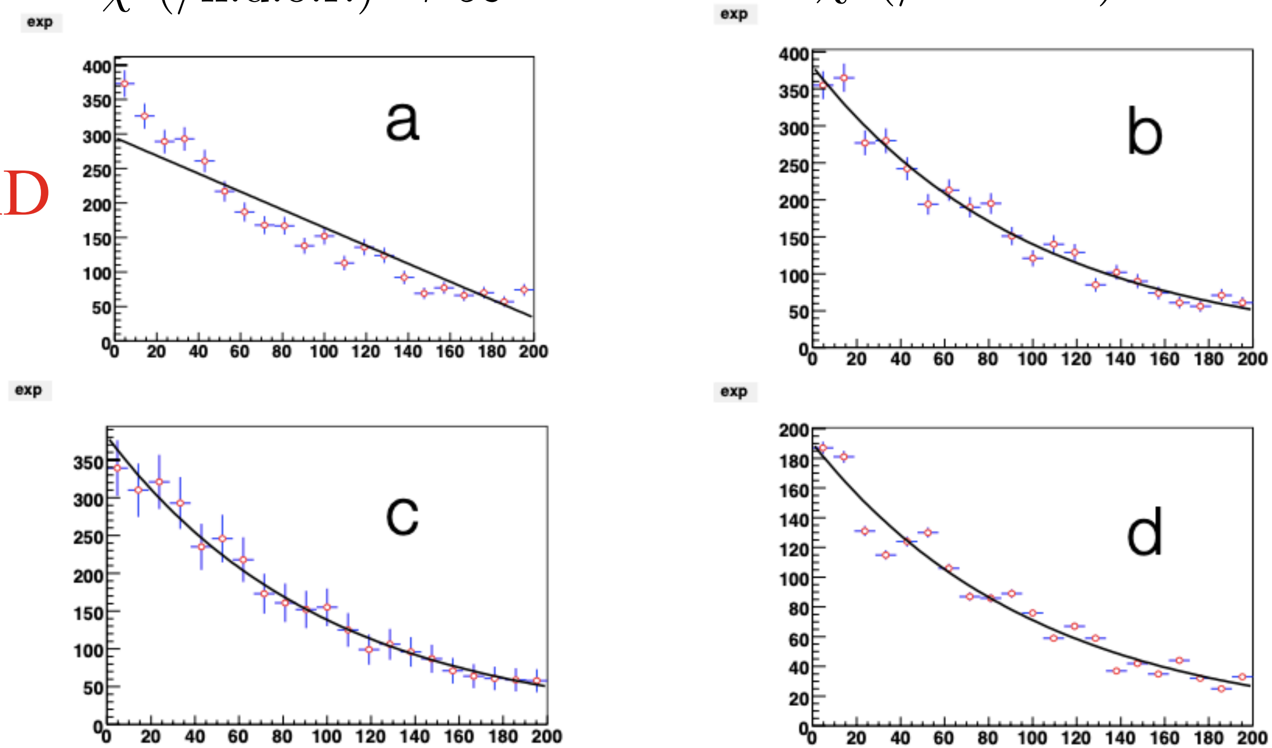
$\chi^2(/\text{n.d.o.f.}) \to \infty$

- Not a serious quantitative inference yet, just a check to proceed further!

- **which fit is making sense?**

$\chi^2(/\text{n.d.o.f.}) \to \infty$ BAD

$\chi^2(/\text{n.d.o.f.}) \sim 1$ GOOD



$\chi^2(/\text{n.d.o.f.}) \sim 0$

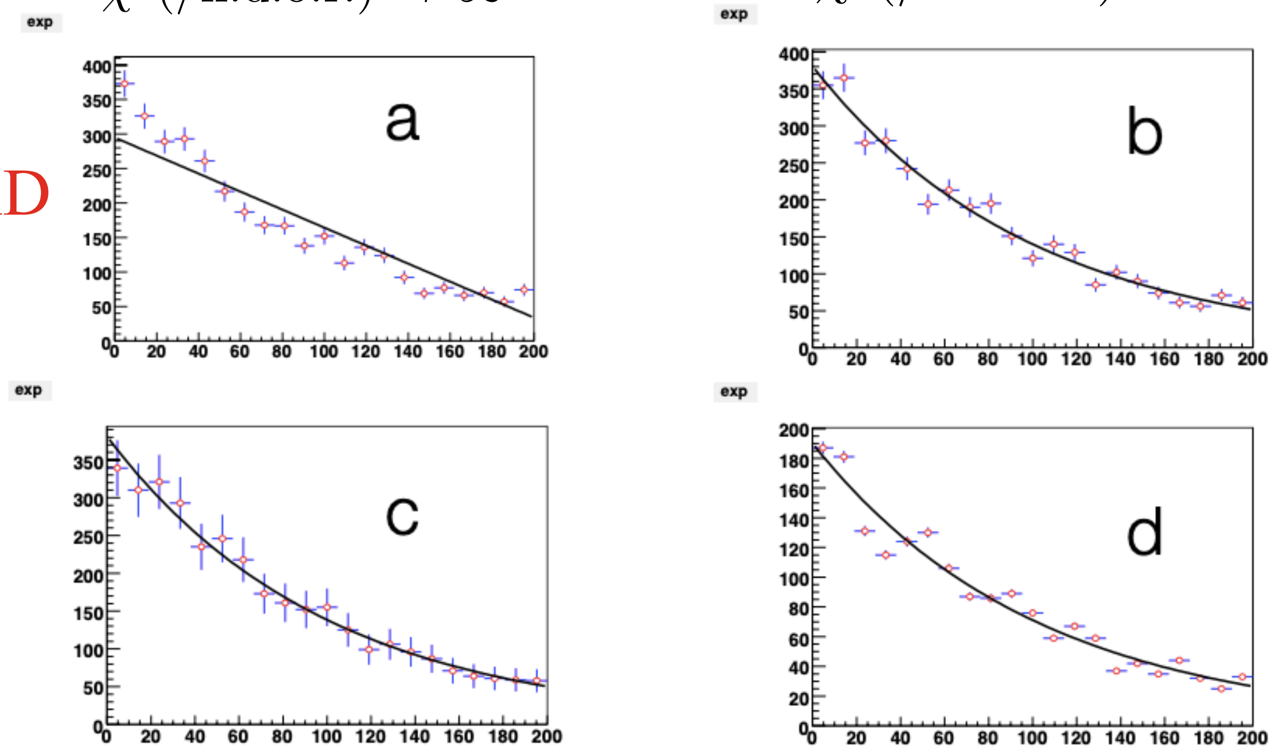
$\chi^2(/\text{n.d.o.f.}) \to \infty$

- Not a serious quantitative inference yet, just a check to proceed further!

- **which fit is making sense?**

$\chi^2(/\text{n.d.o.f.}) \to \infty$ — BAD (a)

$\chi^2(/\text{n.d.o.f.}) \sim 1$ — GOOD (b)



$\chi^2(/\text{n.d.o.f.}) \sim 0$ (c)

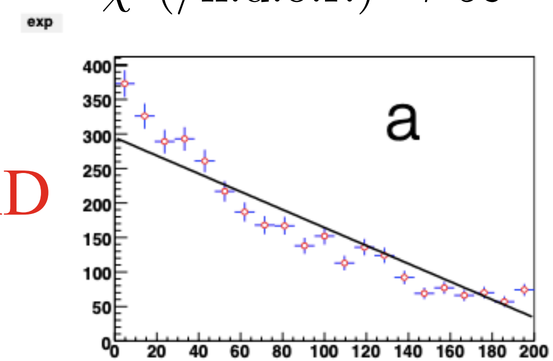
$\chi^2(/\text{n.d.o.f.}) \to \infty$ — BAD (d)

- Not a serious quantitative inference yet, just a check to proceed further!

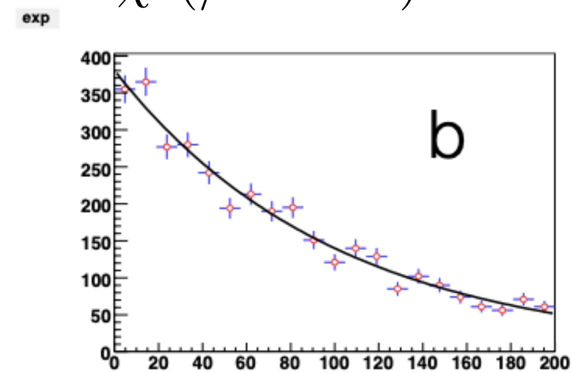
- **which fit is making sense?**

$\chi^2(/\text{n.d.o.f.}) \to \infty$

BAD

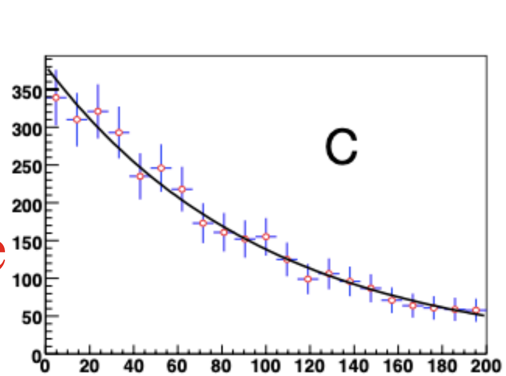


$\chi^2(/\text{n.d.o.f.}) \sim 1$

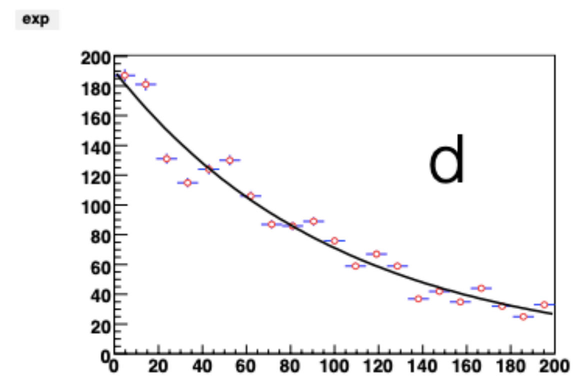


GOOD

LIAR!
or too
conservative
uncert.



$\chi^2(/\text{n.d.o.f.}) \sim 0$

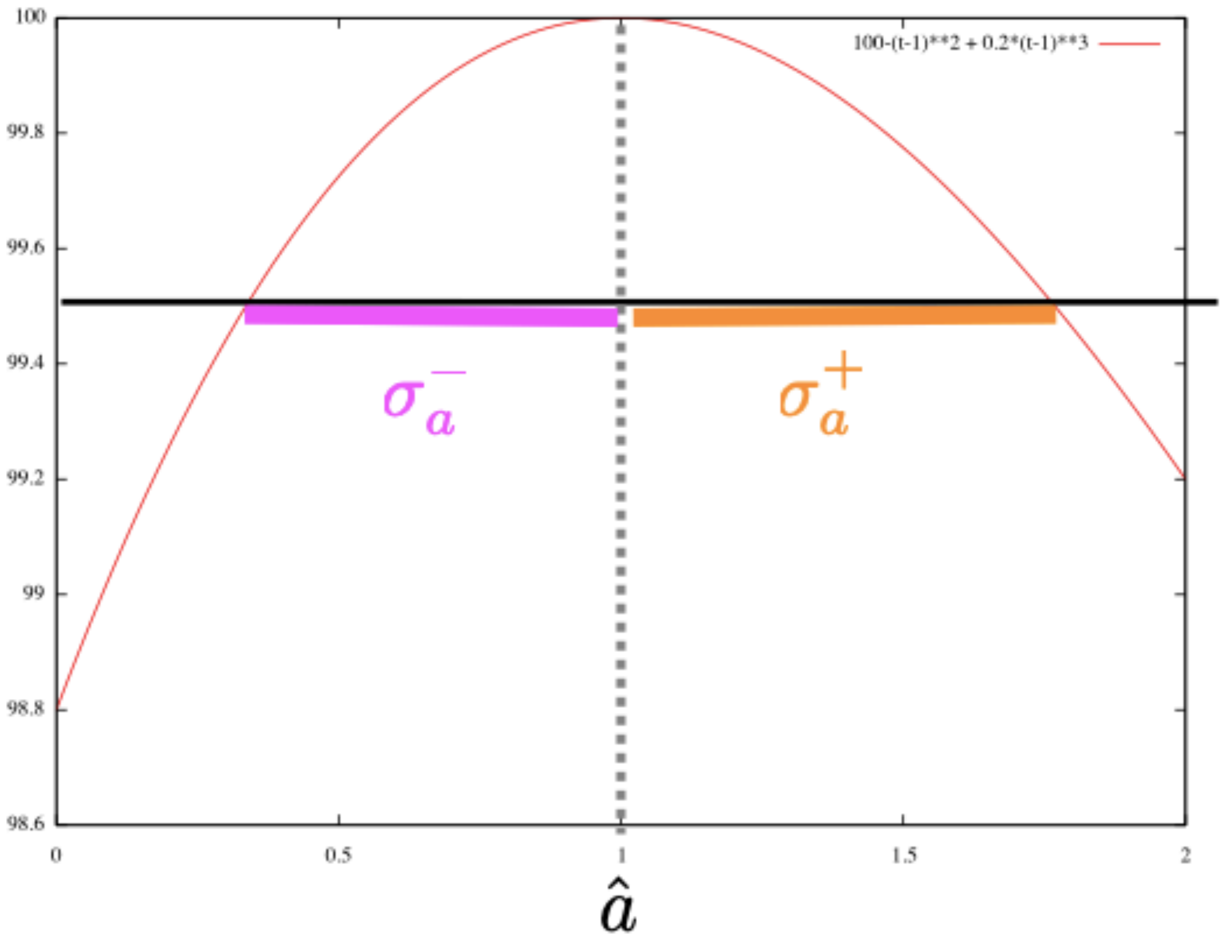


BAD

$\chi^2(/\text{n.d.o.f.}) \to \infty$

- Not a serious quantitative inference yet, just a check to proceed further!

$$a = \hat{a}^{+\sigma_a^+}_{-\sigma_a^-}$$



$$\Delta \ln \mathcal{L} = \frac{1}{2}$$

Chapter IV: Statistical method (Inferences)

The different interpretations of probability

Probabilities can be interpreted in different ways:

- **Interpretation 1: probability = frequency of occurence**

$$P(A) = \lim_{n\to\infty} \frac{n(A)}{n} \quad \text{where} \quad \begin{cases} A\text{: some event} \\ n(A)\text{: number of times event } A \text{ occurs} \\ n\text{: total number of events} \end{cases}$$

Example: dice roll

- **Interpretation 2: probability = degree of belief on occurence**
  - Often used for unique events (for which frequentist calculation impossible)
    - → Example: probability that it'll rain tomorrow
  - Bayesians use it always (even for non-unique events)

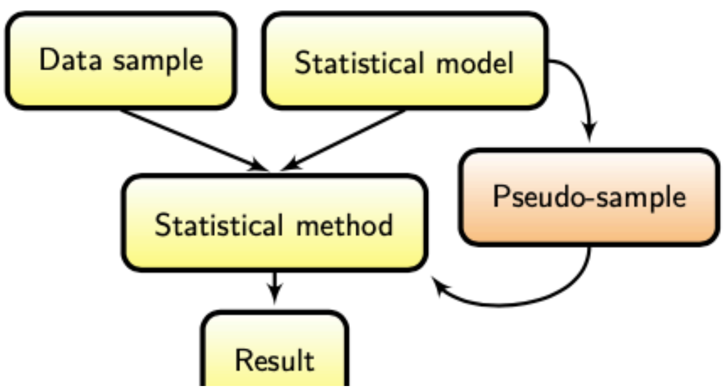
Chapter IV: Statistical method: two paradigms

Frequentist

- Use frequentist interpretation of probability
- Model parameters are fixed, the only thing that can have a probability associated to it is the data

$$P(x;\theta)$$

- Inference based on notion of **pseudo-sample** or **pseudo-dataset**



Bayesian

- Use probabilities as degree of belief
- Model parameters have a probability

$$P(\theta;x)$$

  → They are not random variable as the x's (better qualified as **uncertain variables**)
  → Even if notation is the same, this probability doesn't have the same frequentist meaning

- Inference of parameter of interest based on this probability
- How is this probability computed ?

$$P(\theta;x) = \frac{P(x;\theta)P(\theta)}{P(x)} \quad \text{(Bayes thm)}$$

Chapter IV: Statistical method: two paradigms

- **Some rapid comments about frequentism / bayesianism**

- **Frequentism:** following a fit maximising the likelihood

$$\mathcal{L}(\text{theory}) \equiv \prod_{\text{all data points}} P(\text{datapoint}_i|\text{theory})$$

This is nothing less than **the probability to see the set of data we've measured, given the theory.**

- Note that one would like to get the probability of the theory given the data, the model is the natural one? **That would be the Bayesian probability.**

- Yet, the result is plagued by the hypothetisation of the prior distribution for the parameter(s) of interest. The good news is that its importance decreases when the precision of the data increases.

- Frequentism is on the contrary well-motivated. Very demanding though!

Summaries:

- We touched in this introduction:
  - Some basics of probability law
  - How to address the properties of datasets, be they empirical or through probability laws.
  - We examined the central limit theorem consequences . Repeating an experiment provides you with the gaussian blur.
  - We discussed how make an inference.
  - Most of your statistical problems have already a well-defined solution. We shall always remember that it is mathematics, hence axiomatic!