Probability and Statistics

Basic concepts

(from a physicist point of view)

Benoit CLEMENT – Université J. Fourier / LPSC

bclement@lpsc.in2p3.fr

School Of Statistics

Bibliography

Kendall's Advanced theory of statistics, Hodder Arnold Pub.

volume 1 : Distribution theory, A. Stuart et K. Ord

volume 2a : Classical Inference and and the Linear Model, A. Stuart, K. Ord, S. Arnold

volume 2b : Bayesian inference, A. O'Hagan, J. Forster

The Review of Particle Physics, K. Nakamura et al., J. Phys. G 37, 075021 (2010) (+Booklet)

Data Analysis: A Bayesian Tutorial, D. Sivia and J. Skilling, Oxford Science Publication

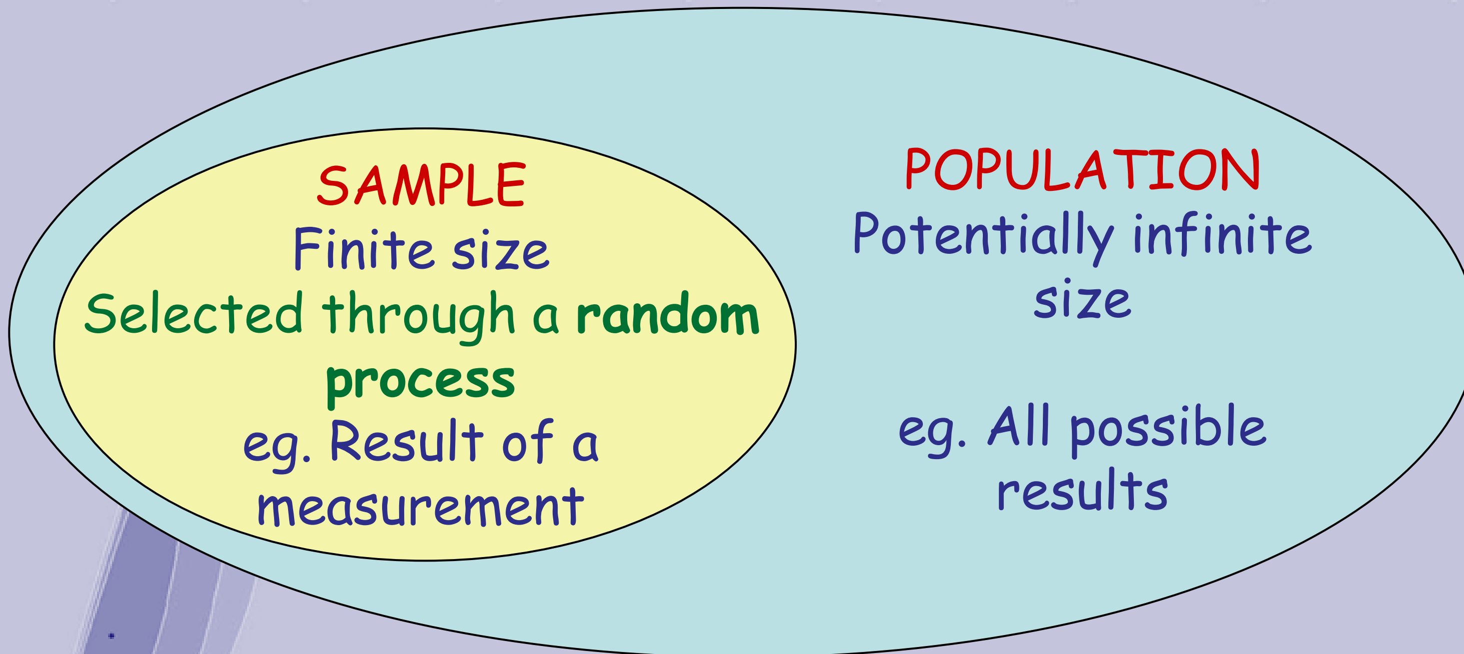
Statistical Data Analysis, Glen Cowan, Oxford Science Publication

Analyse statistique des données expérimentales, K. Protassov, EDP sciences

Probabilités, analyse des données et statistiques, G. Saporta, Technip

Analyse de données en sciences expérimentales, B.C., Dunod

Sample and population



Characterizing the sample, the population and the drawing procedure :

-> **Probability theory** (today's lecture)

Using the sample to estimate the characteristics of the population

-> **Statistical inference** (tomorrow's lecture)

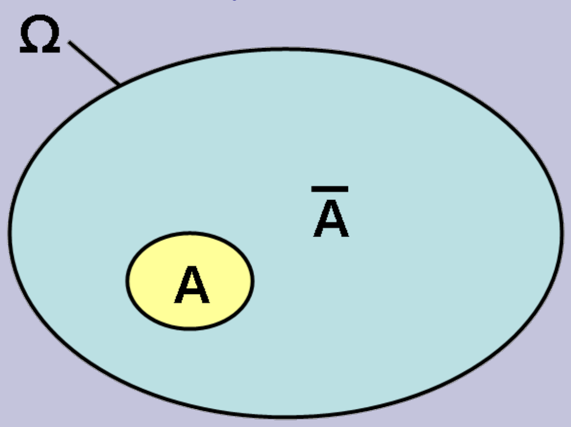
Random process

A random process (« measurement » or « experiment ») is a process whose **outcome cannot be predicted with certainty.**

It will be described by :

Universe: **Ω** = set of all possible **outcomes**.

Event : **logical condition on an outcome**. It can either be true or false; an event splits the universe in 2 subsets.



An event *A* will be identified by the subset **A** for which *A* is true.

Probability

A **probability function** P is defined by : (Kolmogorov, 1933)

$$P : \{\text{Events}\} \rightarrow [0:1]$$
$$A \rightarrow P(A)$$

satisfying :

$$P(\Omega)=1$$
$$P(A \cup B) = P(A) + P(B) \quad \text{if } A \cap B = \varnothing$$

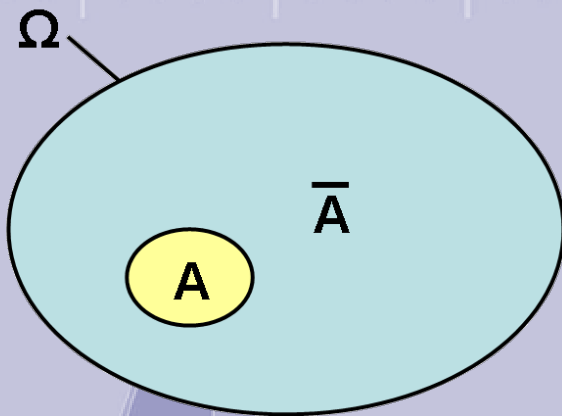
Interpretation of this number :

- **Frequentist approach** : if we repeat the random process a great number of times n , and count the number of times the outcome satisfy event A, n_A then the ratio :

$$\lim_{n \to +\infty} \frac{n_A}{n} = P(A)$$ defines a probability

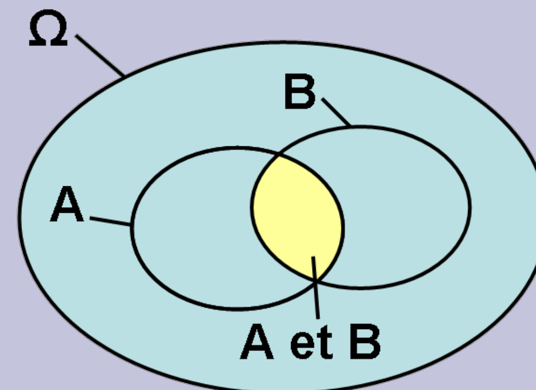
- **Bayesian interpretation** : a probability is a measure of the credibility associated to the event.

Simple logic

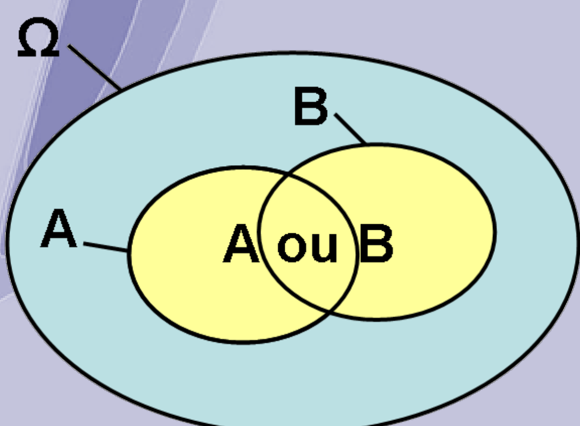


Event « **not** $\mathcal{A}$ » is associated with the complement $\bar{A}$.

$$P(\bar{A}) = 1 - P(A)$$
$$P(\emptyset) = 1 - P(\Omega) = 0$$



Event « $\mathcal{A}$ **and** $\mathcal{B}$ » is associated with the ensemble $A \cap B$.



Event « $\mathcal{A}$ **or** $\mathcal{B}$ » is associated with the ensemble $A \cup B$.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

We will use and mix notations : $\mathcal{A} \leftrightarrow A$, **or** $\leftrightarrow \cup$, **and** $\leftrightarrow \cap$

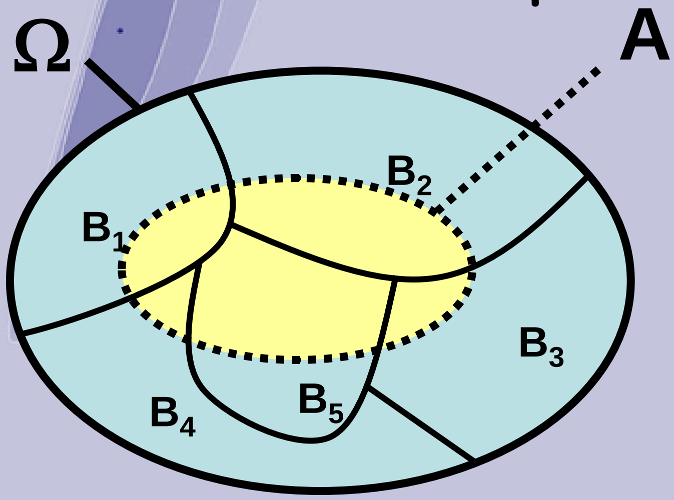
Incompatibility and partition

Two **incompatible** events cannot be true simultaneously, then :

$$P(A \text{ and } B) = 0 \quad \text{and} \quad P(A \text{ or } B) = P(A)+P(B)$$

A **partition** is a set of incompatible events that cover the full universe :

$$\Omega = \bigcup_i B_i, \qquad B_i \cap B_j = \varnothing \quad (i \neq j)$$



Then, for any event *A* :

$$P(A) = \sum_i P(A \text{ and } B_i)$$

Similar to a basis in linear algebra.

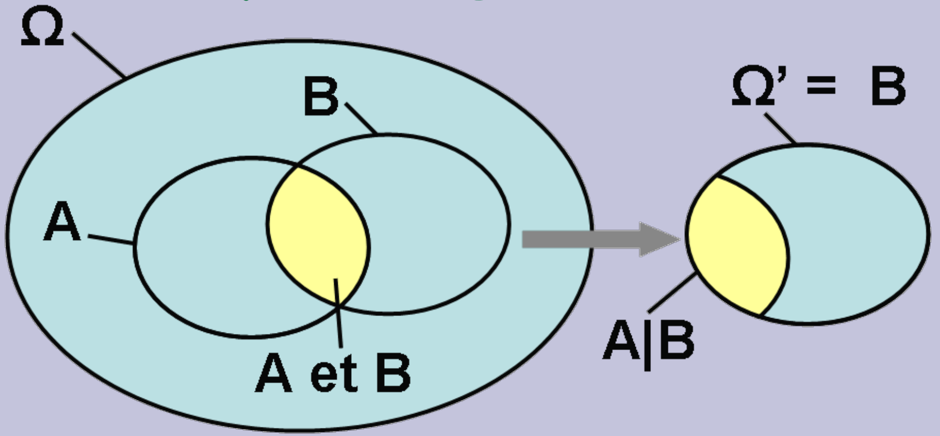
Conditional probability and independence

If an event **B** is **known to be true**, one can restrain the universe to Ω'=B and define a new probability function on this universe, the **conditional probability**.

P(A|B) = « probability of A given B »

From Venn diagram :

$$P(A\,|\,B) = \frac{P(A \text{ and } B)}{P(B)}$$



Two events are **independent**, if the realization of one is not linked in any way to the realization of the other :

P(A|B)=P(A) and P(B|A) = P(B)

From the previous relations: **P(A and B) = P(A).P(B)**

Bayes theorem

The definition of conditional probability leads to :

$P(A \text{ and } B) = P(A|B).P(B) = P(B|A).P(A)$

Hence relating $P(A|B)$ to $P(B|A)$ by the **Bayes theorem :**

$$P(B|A) = \frac{P(A|B).P(B)}{P(A)}$$

Or, using a partition $\{B_i\}$:

$$P(B_i|A) = \frac{P(A|B_i).P(B_i)}{\sum_i P(A \text{ and } B_i)} = \frac{P(A|B_i).P(B_i)}{\sum_i P(A|B_i).P(B_i)}$$

This theorem will play a major role in Bayesian inference : given **data** and a **set of models**, it translates into :

$$P(\text{model}_i|\text{data}) = \frac{P(\text{data}|\text{model}_i).P(\text{model}_i)}{\sum_i P(\text{data}|\text{model}_i).P(\text{model}_i)}$$

Application of Bayes

100 dices in a box :

70 are equiprobable (A) 30 have a probability 1/3 to get 6 (B)

You pick one dice, throw it until you reach 6 and count the number of try. Repeating the process thrice, you get 2, 4 and 1.

What's the probability that the dice is equilibrated ?

For one throw : $P(n|A) = (1-p_6)^{n-1}p_6 = \frac{5^{n-1}}{6^n}$ $P(n|B) = \frac{2^{n-1}}{3^n}$

Combining several throw: (for one dice, throws are independents)

$$P(n_1 \text{ and } n_2 \text{ and } n_3 | A) = P(n_1|A)P(n_2|A)P(n_3|A) = \frac{5^{n_1+n_2+n_3-3}}{6^{n_1+n_2+n_3}}$$

$$P(n_1 \text{ and } n_2 \text{ and } n_3 | B) = \frac{2^{n_1+n_2+n_3-3}}{3^{n_1+n_2+n_3}}$$

$$P(A|n_1, n_2, n_3) = \frac{P(n_1, n_2, n_3|A)P(A)}{P(n_1, n_2, n_3|B)P(B) + P(n_1, n_2, n_3|A)P(A)}$$

$$= \frac{\frac{5^{n_1+n_2+n_3-3}}{6^{n_1+n_2+n_3}} \times 0.7}{\frac{2^{n_1+n_2+n_3-3}}{3^{n_1+n_2+n_3}} \times 0.3 + \frac{5^{n_1+n_2+n_3-3}}{6^{n_1+n_2+n_3}} \times 0.7} = \frac{\frac{5^4}{6^7} \times 0.7}{\frac{2^4}{3^7} \times 0.3 + \frac{5^4}{6^7} \times 0.7} \approx 0.42$$

Random variable

When the outcome of the random process is a **number** (real or integer), we associate to the random process, a **random variable X**.

Each realization of the process leads to a particular result : **X=x**. **x** is a realization of **X**.

For a **discrete variable** :

Probability law : $p(x) = P(X=x)$

For a **real variable** : $P(X=x)=0$,

Cumulative density function : $F(x) = P(X<x)$

$dF = F(x+dx)-F(x) = P(X < x+dx) - P(X < x)$

$= P(X < x \text{ or } x < X < x+dx) - P(X < x)$

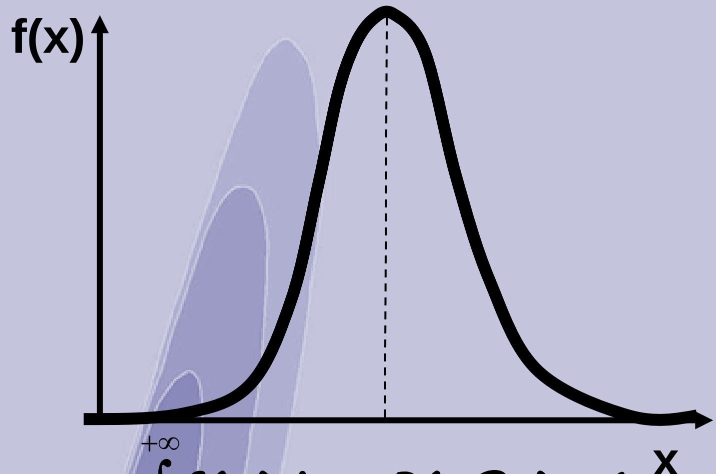
$= P(X < x) + P(x < X < x+dx) - P(X < x)$

$= P(x < X < x+dx) = f(x)dx$

Probability density function (pdf) : $f(x) = \frac{dF}{dx}$

Density function

Probability density function



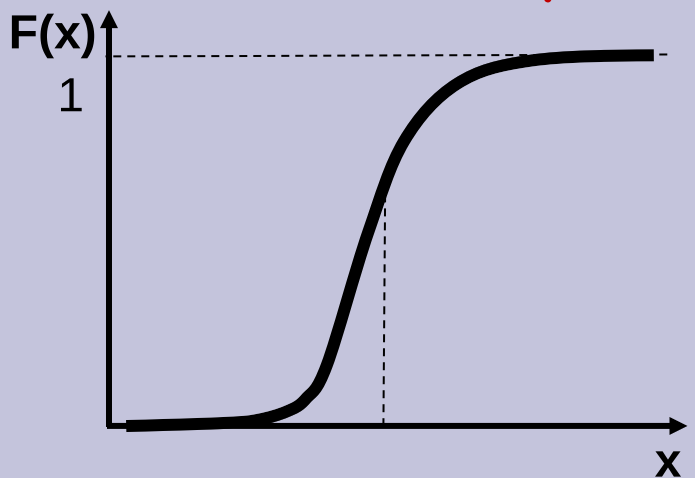
$$\int_{-\infty}^{+\infty} f(x)dx = P(\Omega) = 1$$

Note : discrete variables can also be described by a probability density function using Dirac distributions:

$$f(x) = \sum_i p(i)\delta(i - x)$$

$$\sum_i p(i) = 1$$

Cumulative density function



By construction :

$$F(-\infty) = P(\emptyset) = 0$$

$$F(+\infty) = P(\Omega) = 1$$

$$F(a) = \int_{-\infty}^{a} f(x)dx$$

$$P(a < X < b) = F(b) - F(a) = \int_{a}^{b} f(x)dx$$

Change of variable

Probability density function of $Y = \varphi(X)$

For φ bijective

- φ increasing : $X<x \Leftrightarrow Y<y$

$$P(X<x) = F_X(x) = P(Y<y) = F_y(Y) = F_y(\varphi(x)) \Rightarrow f_y(y) = \frac{dF(x)}{dy} = \frac{f(x)}{\varphi'(x)}$$

- φ decreasing : $X<x \Leftrightarrow Y>y$

$$P(X<x) = F_X(x) = P(Y>y) = 1 - F_y(Y) = 1 - F_y(\varphi(x)) \Rightarrow f_y(y) = -\frac{dF(x)}{dy} = \frac{f(x)}{-\varphi'(x)}$$

in both case

$$f_y(y) = \frac{f(x)}{|\varphi'(x)|} = \frac{f(\varphi^{-1}(y))}{|\varphi'(\varphi^{-1}(y))|}$$

If φ not bijective : split into several bijective parts φ_i

$$f_y(y) = \sum_i \frac{f(x)}{|\varphi_i'(x)|} = \sum_i \frac{f(\varphi_i^{-1}(y))}{|\varphi_i'(\varphi_i^{-1}(y))|}$$

Very useful for Monte-Carlo : if X is uniformly distributed between 0 and 1 then $Y=F^{-1}(X)$ has F for cumulative density

Multidimensional PDF (1)

Random variables can be generalized to random vectors :

$$\vec{X} = (X_1, X_2, \dots, X_n)$$

the **probability density function** becomes :

$$f(\vec{x})d\vec{x} = f(x_1, x_2, \dots, x_n)dx_1 dx_2 \dots dx_n$$
$$= P(x_1 < X_1 < x_1 + dx_1 \text{ and } x_2 < X_2 < x_2 + dx_2 \dots$$
$$\dots \text{and } x_n < X_n < x_n + dx_n)$$

and $$P(a < X < b \text{ and } c < Y < d) = \int_a^b dx \int_c^d dy\ f(x,y)$$

Marginal density : probability of only one of the component

$$f_X(x)dx = P(x < X < x + dx \text{ and } -\infty < Y < +\infty) = \int \left(f(x,y)dx\right)dy$$

$$\Rightarrow f_X(x) = \int f(x,y)dy$$

Multidimensional PDF (2)

For a fixed value of $Y=y_0$:

$f(x|y_0)dx$ = « Probability of $x<X<x+dx$ knowing that $Y=y0$ »

is , a **conditional density** for X. It is proportional to $f(x,y)$, so

$$f(x|y) \propto f(x,y) \qquad \int f(x|y)dx = 1$$

$$\Rightarrow f(x|y) = \frac{f(x,y)}{\int f(x,y)dx} = \frac{f(x,y)}{f_Y(y)}$$

The two random variables X and Y are **independent** if all events of the form **$x<X<x+dx$** are independent from **$y<Y<y+dy$**

$f(x|y)=f_X(x)$ and $f(y|x)=f_Y(y)$ hence $f(x,y)= f_X(x).f_Y(y)$

Translated in term of pdf's, Bayes' theorem becomes:

$$f(y|x) = \frac{f(x|y)f_Y(y)}{f_X(x)} = \frac{f(x|y)f_Y(y)}{\int f(x|y)f_Y(y)dy}$$

D.Sivia's lecture will detail the use of this formula for statistical inference

Sample PDF

A **sample** is obtained from a **random drawing** within a **population**, described by a probability density function.

We're going to discuss how to **characterize, independently from one another:**

- a **population**
- a **sample**

To this end, it is useful, to consider a sample as a finite set from which one can randomly draw elements, with equipropability

We can the associate to this process a probability density, the **empirical density** or **sample density**

$$f_{sample}(x) = \frac{1}{n}\sum_{i}\delta(x-i)$$

This density will be useful to translate properties of distribution to a finite sample.

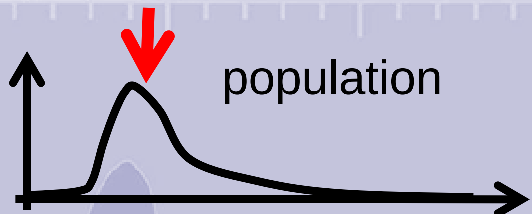
Characterizing a distribution

How to reduce a distribution/sample to a finite number of values ?

- **Measure of location**:
  Reducing the distribution to **one central value**
  -> **Result**
- **Measure of dispersion**:
  Spread of the distribution around the central value
  -> **Uncertainty/Error**
- **Higher order measure of shape**

- **Frequency table/histogram** (for a finite sample)

Measure of location



Mean value : Sum (integral) of all possible values weighted by the probability of occurrence:

$$\mu = \overline{x} = \int_{-\infty}^{+\infty} x f(x) dx \qquad \mu = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Median : Value that split the distribution in 2 equiprobable parts

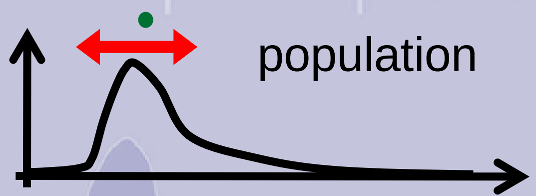
$$x_1 \le x_2 \le \ldots \le x_n$$

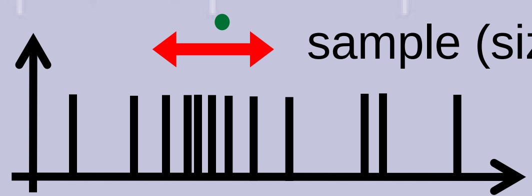
$$\int_{-\infty}^{med(x)} f(x)dx = \int_{med(x)}^{+\infty} f(x)dx = \frac{1}{2} \qquad med(x) = \begin{cases} x_{(n+1)/2} & , \quad \text{odd } n \\ \frac{1}{2}\left(x_{n/2} + x_{1+n/2}\right), & \text{even } n \end{cases}$$

Mode : The most probable value = maximum of pdf

$$\left.\frac{df}{dx}\right|_{x=mod(x)} = 0, \quad \left.\frac{d^2 f}{dx^2}\right|_{x=mod(x)} < 0 \qquad ?$$

Measure of dispersion





Standard deviation (σ) and variance (v= σ²) : Mean value of the squared deviation to the mean :

$$v = \sigma^2 = \int (x-\mu)^2 f(x)dx \qquad v = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

Koenig's theorem :

$$\sigma^2 = \int x^2 f(x)dx + \mu^2 \int f(x)dx - 2\mu \int x f(x)dx$$

$$\boxed{\sigma^2 = \overline{x^2} - \mu^2 = \overline{x^2} - \overline{x}^2}$$

Interquartile difference : generalize the mean by splitting the distribution in 4 :

$$\int_{-\infty}^{q_1} f(x)dx = \int_{q_1}^{q_2} f(x)dx = \int_{q_2}^{q_3} f(x)dx = \int_{q_3}^{+\infty} f(x)dx = \frac{1}{4} \qquad \begin{aligned} med(x) &= q_2 \\ \delta &= q_3 - q_1 \end{aligned}$$

Others…

Bienaymé-Chebyshev

Consider the interval : $\Delta =]-\infty, \mu - a[\cup]\mu + a, +\infty[$

Then for $x \in \Delta$: $\left(\frac{x-\mu}{a}\right)^2 > 1 \Rightarrow \left(\frac{x-\mu}{a}\right)^2 f(x) > f(x)$

$$\Rightarrow \int_\Delta \left(\frac{x-\mu}{a}\right)^2 f(x)dx > \int_\Delta f(x)dx$$

$$\Rightarrow \int_{-\infty}^{+\infty} \left(\frac{x-\mu}{a}\right)^2 f(x)dx > \int_\Delta f(x)$$

$$\Rightarrow \frac{\sigma^2}{a^2} > P(|X-\mu| > a)$$

Finally **Bienaymé-Chebyshev's inequality** $P(|X-\mu| \le a\sigma) > 1 - \frac{1}{a^2}$

It gives a bound on the **confidence level** if the interval $\mu \pm a\sigma$

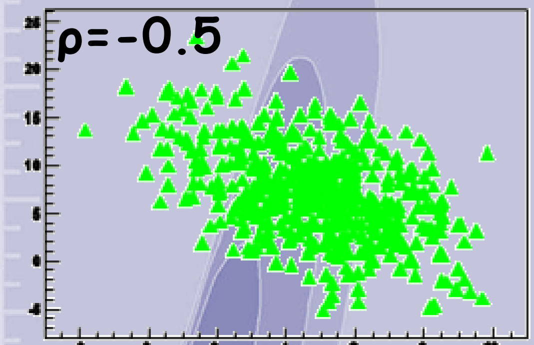
a	1	2	3	4	5
Chebyshev's bound	0	0.75	0.889	0.938	0.96
Normal distribution	0.683	0.954	0.997	0.99996	0.9999994

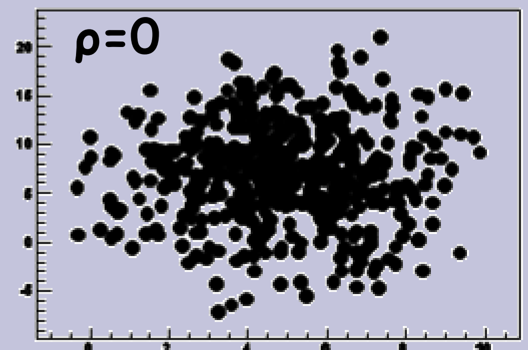
Multidimensional case

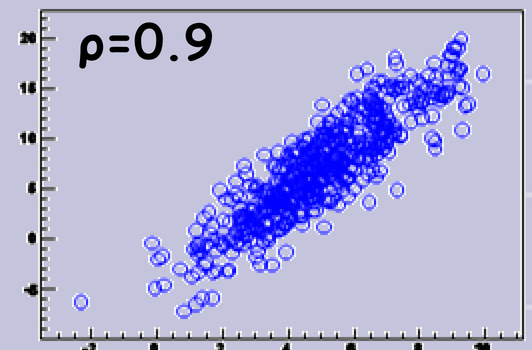
A random vector (X,Y) can be treated as **2 separate variables**

marginal densities : mean and variance for each variable : $\mu_X\ \mu_Y\ \sigma_X\ \sigma_Y$

Doesn't take into account correlations between the variables





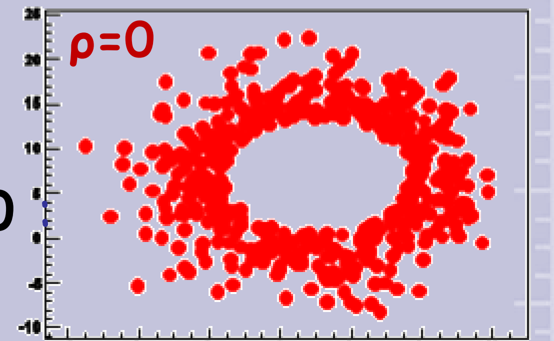


Generalized measure of dispersion : **Covariance of X and Y**

$$Cov(X,Y) = \iint (x-\mu_X)(y-\mu_Y) f(x,y)\,dx\,dy = \rho\sigma_X\sigma_Y = \mu_{XY} - \mu_X\mu_Y$$

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_X)(y_i-\mu_Y)$$

Correlation : $\rho = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}$ **Uncorrelated** : $\rho=0$



Independent ⇄ **Uncorrelated**

only quantify linear correlation

Decorrelation

Covariance matrix for **n** variables $\mathbf{X}_i$:

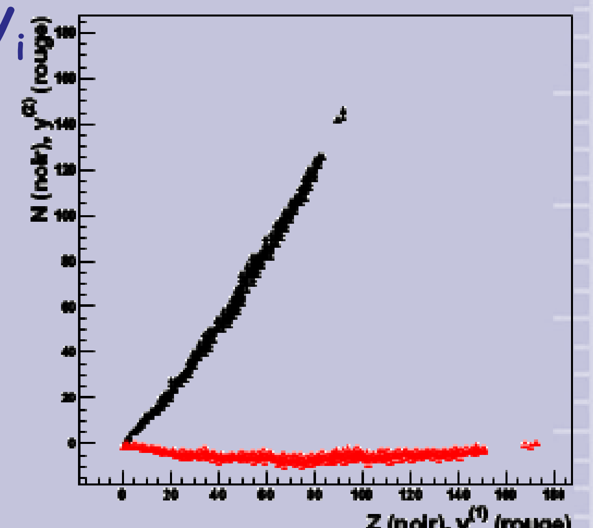
$$\Sigma_{ij} = \mathrm{Cov}(X_i, X_j) \Rightarrow \Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n}\sigma_1\sigma_n & \rho_{2n}\sigma_2\sigma_n & \cdots & \sigma_n^2 \end{pmatrix}$$

For **uncorrelated variables** Σ is **diagonal**

Matrix **real** and **symmetric** **:** can de diagonalized

On can define **n** new uncorrelated variables $\mathbf{Y}_i$

$$\Sigma' = \begin{pmatrix} \sigma_1'^2 & 0 & \cdots & 0 \\ 0 & \sigma_2'^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n'^2 \end{pmatrix} = B^{-1}\Sigma B, \qquad Y = BX$$



$\sigma_i'^2$ are the **eigenvalues** of Σ,
B contains the **orthonormal eigenvectors**.

The $\mathbf{Y}_i$ are the **principal components**. Sorted for the larger to the smaller σ' they allow **dimensionary reduction**

Regression

Measure of location:

- a point : (μ_X , μ_Y)
- a curve : line closest to the points -> **linear regression**

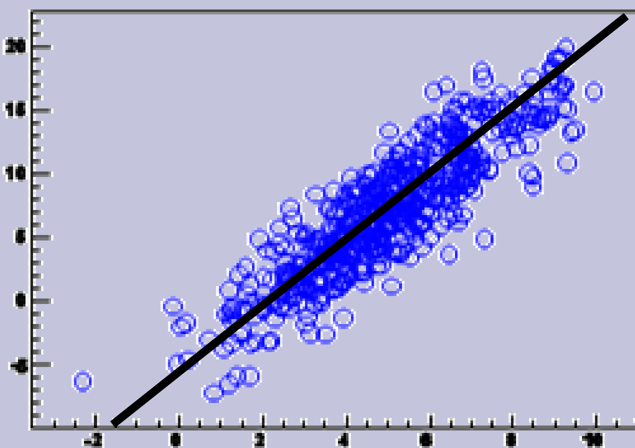
Minimizing the dispersion between the curve « y=ax+b » and the distribution :

$$w(a,b) = \iint (y - ax - b)^2 f(x,y)dxdy \left(= \frac{1}{n} \sum_i (y_i - ax_i - b)^2 \right)$$

$$\begin{cases} \dfrac{\partial w}{\partial a} = 0 = \iint x(y - ax - b) f(x,y)dxdy \\ \dfrac{\partial w}{\partial b} = 0 = \iint (y - ax - b) f(x,y)dxdy \end{cases}$$

$$\Leftrightarrow \begin{cases} a(\sigma_X^2 - \mu_X^2) + b\mu_X = \rho\sigma_X\sigma_Y + \mu_X\mu_Y \\ a\mu_X + b = \mu_Y \end{cases}$$

$$\Leftrightarrow \begin{cases} a = \rho \dfrac{\sigma_Y}{\sigma_X} \\ b = \mu_Y - \rho \dfrac{\sigma_Y}{\sigma_X} \mu_X \end{cases}$$



Fully correlated ρ=1

Fully anti-correlated ρ=-1

Then Y = aX+b

Moments

For any function **g(x)**, the **expectation** of **g** is :

$$E[g(X)] = \int g(x) f(x) dx$$ It's the mean value of g

Moments μ_k are the expectation of X^k.

0th moment : $\mu_0 = 1$ (pdf normalization)

1st moment : $\mu_1 = \mu$ (mean)

$X' = X - \mu_1$ is a **central variable**

2nd central moment : $\mu'_2 = \sigma^2$ (variance)

Characteristic function $\varphi(t) = E[e^{ixt}] = \int f(x) e^{ixt} dx = FT^{-1}[f]$

From Taylor expansion : $\varphi(t) = \int \sum_k \frac{(itx)^k}{k!} f(x) dx = \sum_k \frac{(it)^k}{k!} \mu_k$

$$\mu_k = -i^k \left. \frac{d^k \varphi}{dt^k} \right|_{t=0}$$

Pdf entirely defined by its moments

CF : useful tool for demonstrations

Skewness and kurtosis

Reduced variable : **X'' = (X-μ)/σ = X'/ σ**

Measure of asymmetry :

3rd reduced moment : $\mu''_3 = \sqrt{\beta_1} = \gamma_1$: **skewness**

$\gamma_1 = 0$ for symmetric distribution. Then **mean = median**

Measure of shape :

4th reduced moment : $\mu''_4 = \beta_2 = \gamma_2 + 3$: **kurtosis**

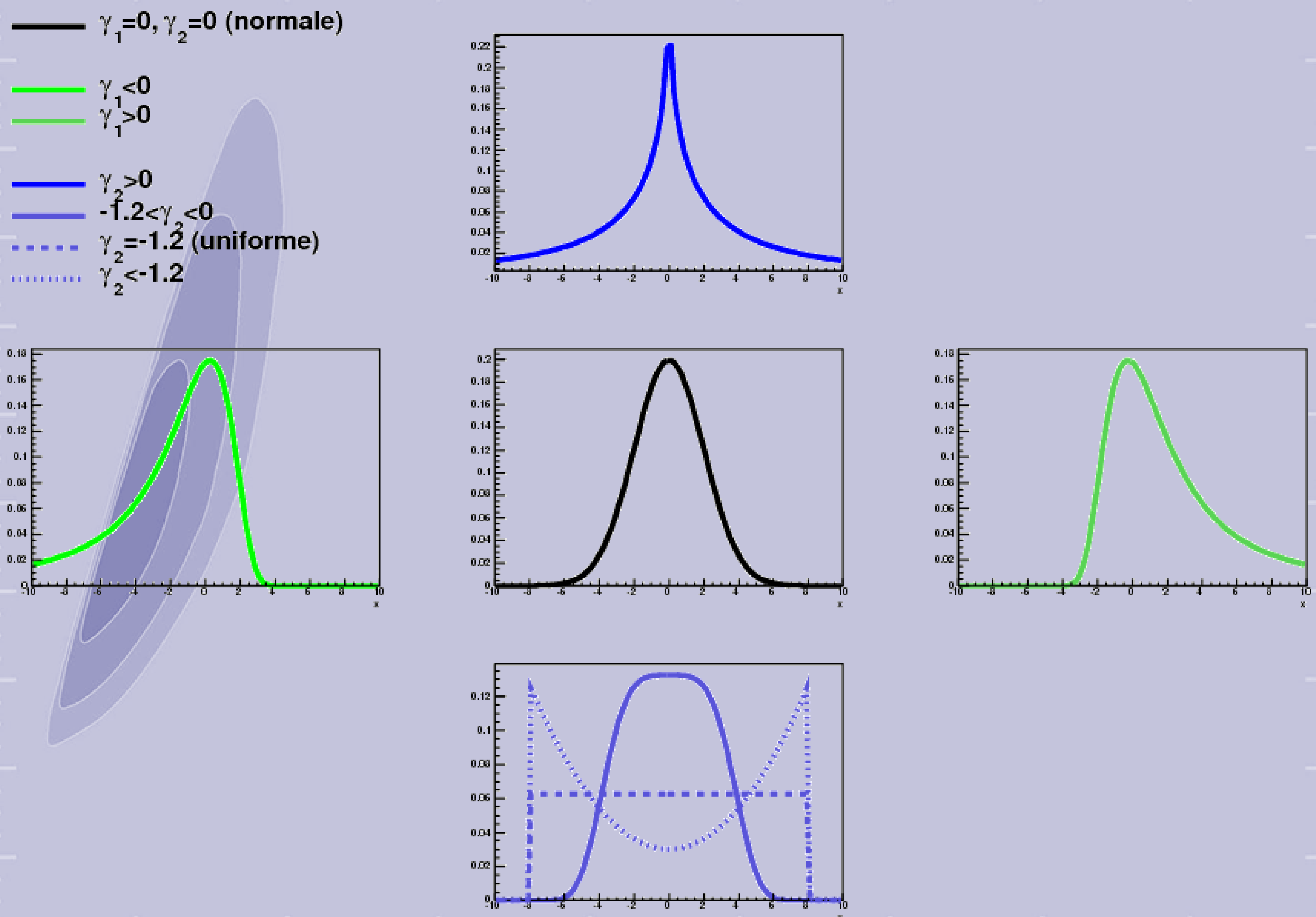
For the normal distribution $\beta_2 = 3$ and $\gamma_2 = 0$

Generalized Koenig's theorem

$$\mu'_n = (-1)^n (1-n)\mu_1^n + \sum_{k=2}^{n} \frac{n!}{k!(n-k)!} (-\mu_1)^{n-k} \mu_k$$

$$\mu''_n = \left(\frac{1}{\mu'_2}\right)^{n-2} \mu'_n$$

Skewness and kurtosis (2)



Discrete distributions

Binomial distribution: randomly choosing **K** objects within a finite set of **n**, with a fixed drawing probability of **p**

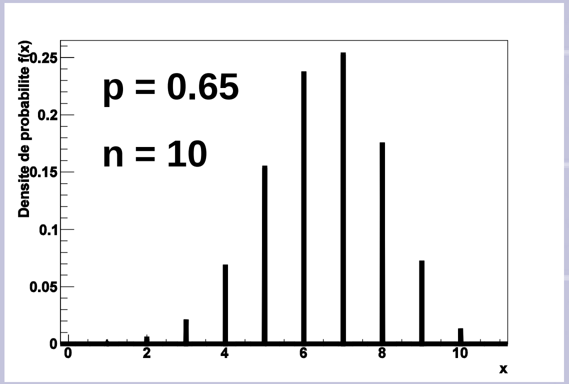
Variable : **K**

Parameters : **n,p**

Law : $P(k;n,p) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$

Mean : **np**

Variance : **np(1-p)**



Poisson distribution : limit of the binomial when $n \to +\infty, p \to 0, np = \lambda$

Counting events with fixed probability per time/space unit.

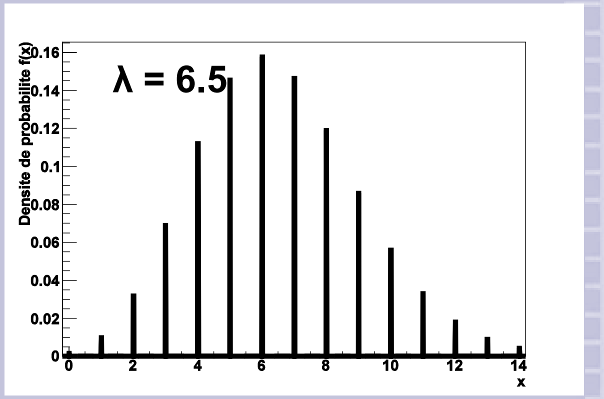
Variable : **K**

Parameters : λ

Law : $P(k;\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$

Mean : λ

Variance : λ



Real distributions

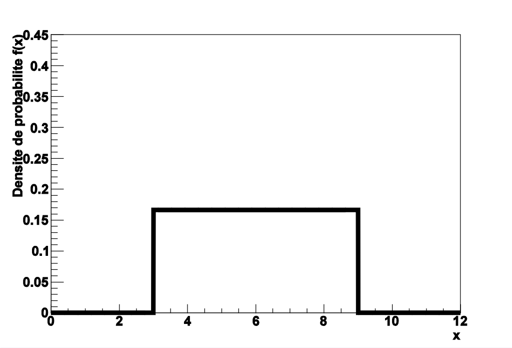
Uniform distribution : equiprobability over a finite range [a,b]

Parameters : a,b

Law : $f(x;a,b) = \frac{1}{b-a}$ if $a < x < b$

Mean : $\mu = (a+b)/2$

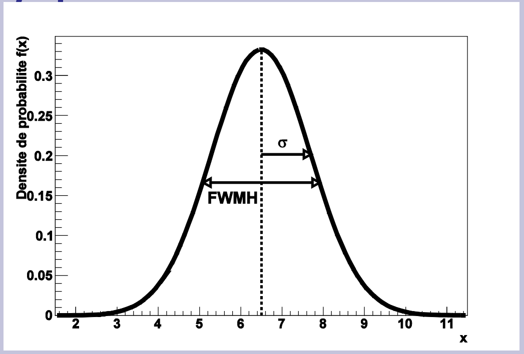
Variance : $v = \sigma^2 = (b-a)^2/12$



Normal distribution (Gaussian) : limit of many processes

Parameters : μ, σ

Law : $f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$



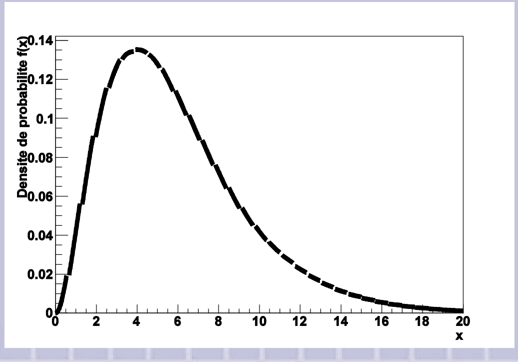
Chi-square distribution : sum of the square of n normal reduced variables

Variable : $c = \sum_{k=1}^{n} \left(\frac{X_k - \mu_{X_k}}{\sigma_{X_k}} \right)^2$

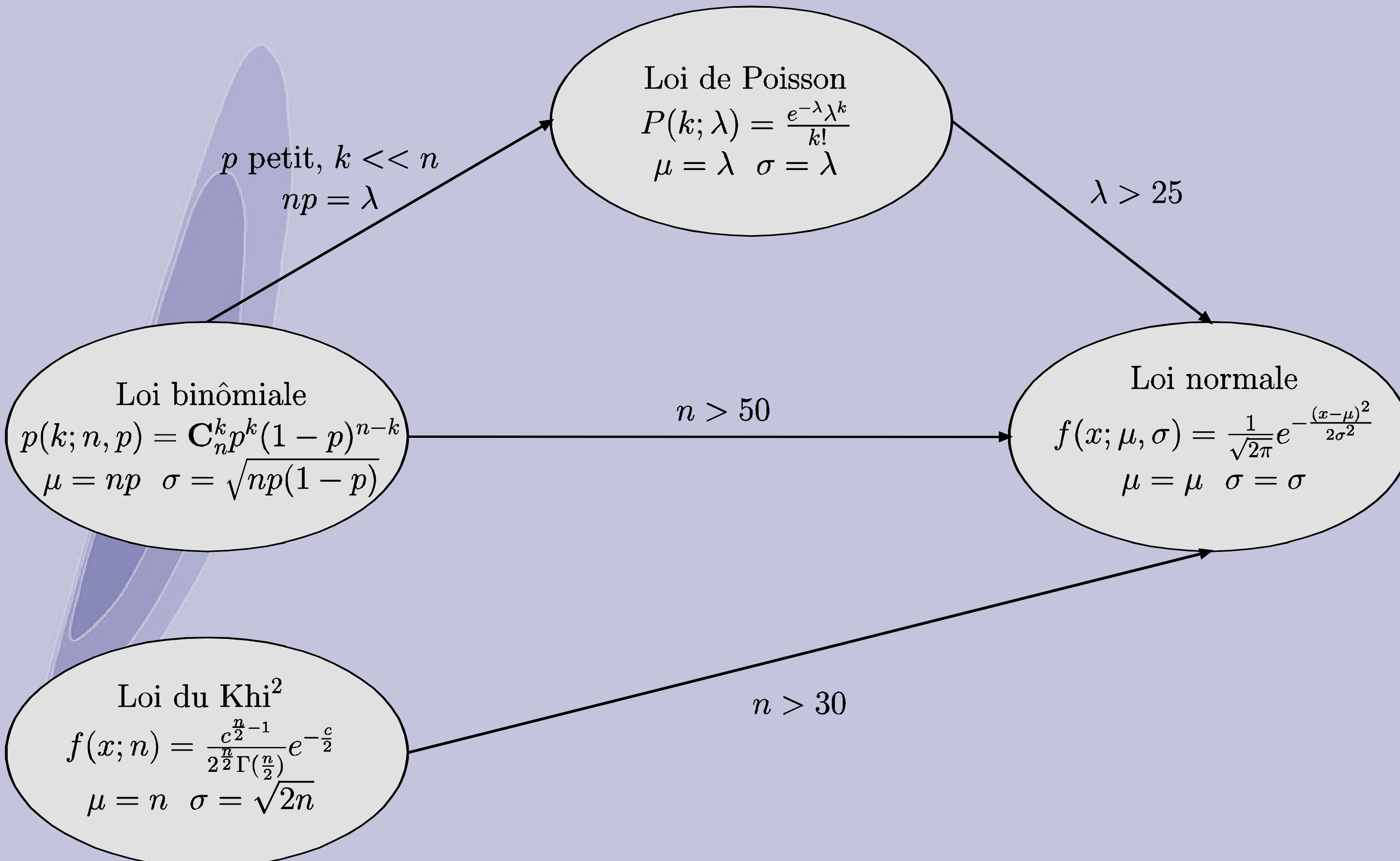
Parameters : n

Law : $f(c;n) = c^{\frac{n}{2}-1} e^{-\frac{c}{2}} \Big/ 2^{\frac{n}{2}-1} \Gamma\left(\frac{n}{2}\right)$

Mean : n Variance : 2n



Convergence



Multidimensional Pdfs

Multinomial distribution : randomly choosing K_1 , K_2 ,… K_s objects within a finite set of **n**, with a fixed drawing probability for each category p_1, p_2,… p_s with $\Sigma K_i=n$ and $\Sigma p_i=1$

Parameters : $n,\ p_1,\ p_2,\dots\ p_s$

Law : $$P(\vec{k};n,\vec{p})=\frac{n!}{k_1!k_2!\dots k_s!}p_1^{k_1}p_2^{k_2}\dots p_s^{k_s}$$

Mean : $\mu_i=np_i$

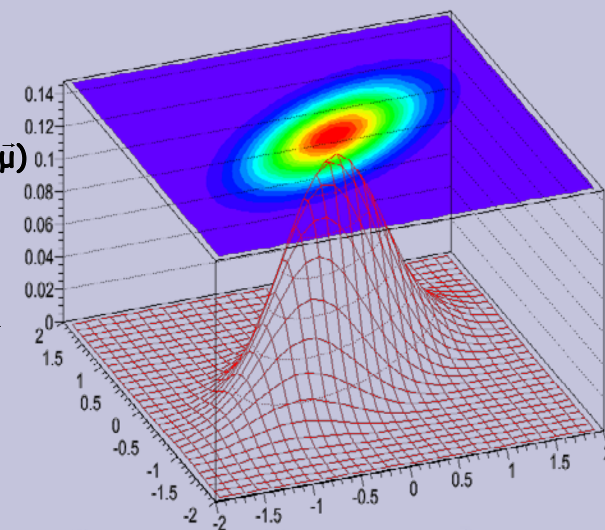
Variance : $\sigma_i^2=np_i(1-p_i)$ $\quad Cov(K_i,K_j)=-np_ip_j$

Rem : **variables are not independent**. The binomial, correspond to s=2, but has only one independent variable.

Multinormal distribution :

Parameters : $\vec{\mu},\Sigma$

Law : $$f(\vec{x};\vec{\mu},\Sigma)=\frac{1}{\sqrt{2\pi|\Sigma|}}e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})}$$



if **uncorrelated** $$f(\vec{x};\vec{\mu},\Sigma)=\prod\frac{1}{\sigma_i\sqrt{2\pi}}e^{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}}$$

Independent ⇄ **Uncorrelated**

Sum of random variables

The sum of several random variable is a new random variable **S**

$$S = \sum_{i=1}^{n} X_i$$

Assuming the mean and variance of each variable exists,

Mean value of S :

$$\mu_S = \int \left(\sum_{i=1}^{n} x_i \right) f(x_1, \ldots, x_n) dx_1 \ldots dx_n = \sum_{i=1}^{n} \int x_i f_{X_i}(x_i) dx_i = \sum_{i=1}^{n} \mu_i$$

The mean is an additive quantity

Variance of S :

$$\sigma_S^2 = \int \left(\sum_{i=1}^{n} x_i - \mu_{X_i} \right)^2 f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

$$= \sum_{i=1}^{n} \sigma_{X_i}^2 + 2 \sum_i \sum_{j<i} Cov(X_i, X_j)$$

$$\sigma_S^2 = \sum_{k=1}^{n} \sigma_{X_i}^2$$

For uncorrelated variables,

the variance is additive -> used for error combinations

Sum of random variables

Probability density function of **S** : $f_S(s)$

Using the characteristic function :

$$\varphi_S(t) = \int f_S(s) e^{ist} ds = \int f_{\vec{X}}(\vec{x}) e^{it\sum x_i} d\vec{x}$$

For **independent variables**

$$\varphi_S(t) = \prod \int f_{X_k}(x_k) e^{itx_k} dx_k = \prod \varphi_{X_i}(t)$$

The characteristic function factorizes.

Finally the pdf is the **Fourier transform** of the cf, so :

$$f_S = f_{X_1} * f_{X_2} * \ldots * f_{X_n}$$

The pdfs of the sum is a **convolution**.

Sum of Normal variables -> Normal

Sum of Poisson variables (λ_1 and λ_2) -> Poisson, $\lambda = \lambda_1 + \lambda_2$

Sum of Khi-2 variables (n_1 and n_2) -> Khi-2, $n = n_1 + n_2$

Sum of independent variables

Weak law of large numbers

Sample of size **n** = realization of **n** independent variables, with the **same distribution** (mean **μ**, variance **σ²**).

The **sample mean** is a realization of $M = \frac{S}{n} = \frac{1}{n}\sum X_i$

Mean value of **M** : $\mu_M = \mu$

Variance of **M** : $\sigma_M^2 = \sigma^2/n$

From Bienaymé-Chebyshev : $P(|M-\mu| > a) < \frac{\sigma^2}{na^2} \xrightarrow[n\to+\infty]{} 0 \ (\forall a)$

Central-Limit theorem

n independent random variables of mean μ_i and variance σ_i^2

Sum of the reduced variables : $C = \frac{1}{\sqrt{n}}\sum \frac{X_i - \mu_i}{\sigma_i}$

The pdfs of **C** converge to a reduced normal distribution :

$$f_C(c) \xrightarrow[n\to+\infty]{} \frac{1}{\sqrt{2\pi}} e^{-\frac{c^2}{2}}$$

Central limit theorem

Naive demonstration:

For each X_i : X''_i has mean 0 and variance 1. So its characteristic function is :

$$\varphi_{X''_i}(t) = 1 - \frac{t^2}{2} + o(t^2)$$

Hence the characteristic function of C :

$$\varphi_C(t) = \varphi_{X''_i}\left(\frac{t}{\sqrt{n}}\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n$$

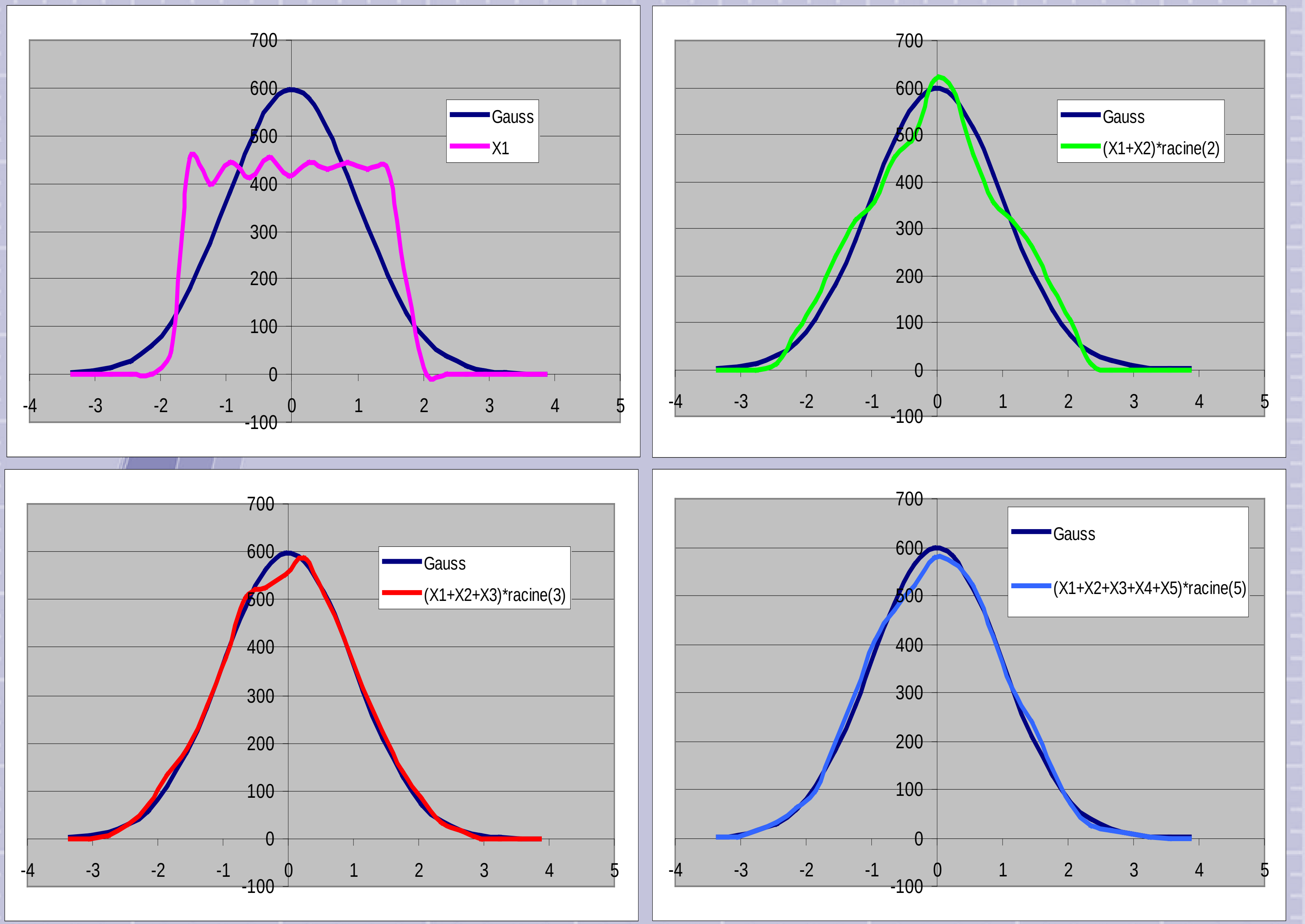
For n large :

$$\lim_{n \to +\infty} \varphi_C(t) = \lim_{n \to +\infty} \left(1 - \frac{t^2}{2n}\right)^n = e^{-\frac{t^2}{2}} = FT^{-1}[f_C]$$

This is a naive demonstration, because we assumed that the moments were defined.

For CLT, only mean and variance are required (much more complex)

Central limit theorem



Dispersion and uncertainty

Any measure (or combination of measure) is a realization of a random variable.

- Measured value : **θ**
- True value : **θ_0**

Uncertainty = quantifying the difference between **θ** and **θ_0** :

->measure of dispersion

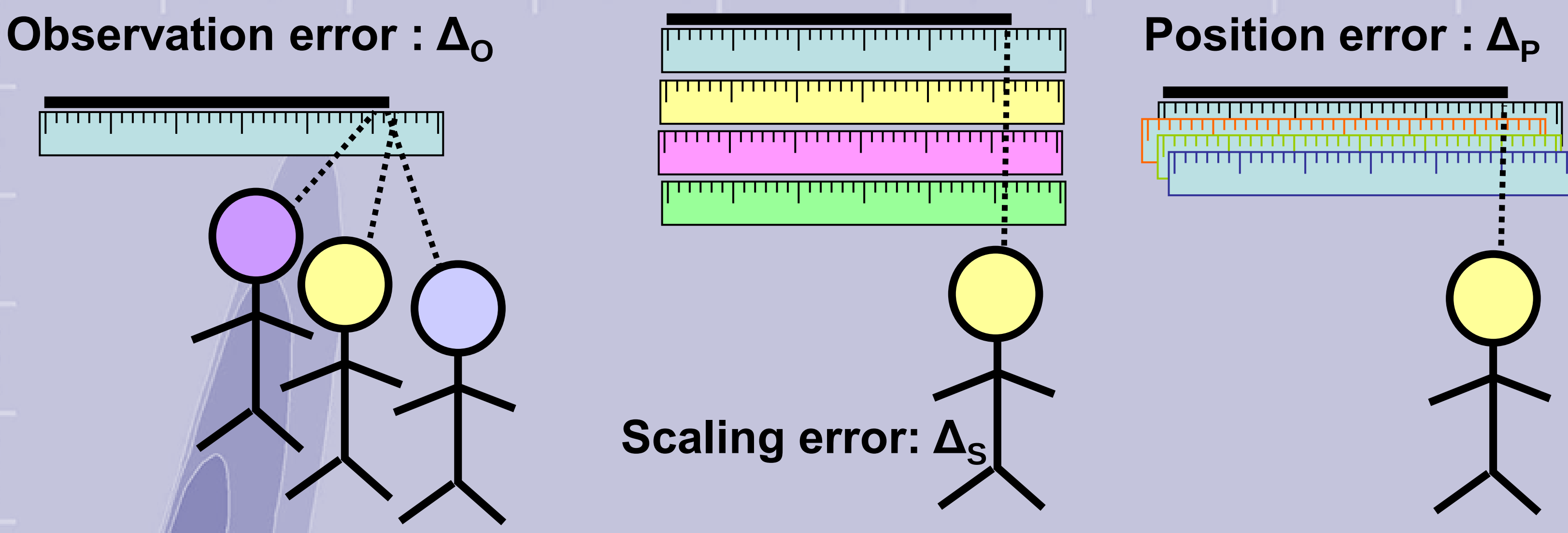
We will postulate : **$\Delta\theta = \alpha\sigma_\theta$** Absolute error, always positive

Usually one differentiate

- **Statistical error** : due to the measurement Pdf.
- **Systematic errors** or bias -> fixed but unknown deviation (equipment, assumptions,…)

Systematic errors can be seen as statistical error in a set a similar experiences.

Error sources



$$\theta = \theta_0 + \delta_O + \delta_S + \delta_P$$

Each δ_i is a realization of a random variable : mean 0 (negligible) and variance σ_i^2. For **uncorrelated error sources :**

$$\left.\begin{aligned} \Delta_O &= \alpha\sigma_O \\ \Delta_S &= \alpha\sigma_S \\ \Delta_P &= \alpha\sigma_P \end{aligned}\right\} \Delta_{tot}^2 = (\alpha\ \sigma_{tot})^2 = \alpha^2(\sigma_O{}^2 + \sigma_S{}^2 + \sigma_P{}^2) = \Delta_O^2 + \Delta_S^2 + \Delta_P^2$$

Choice of α ?

If many sources, from central-limit -> normal distribution

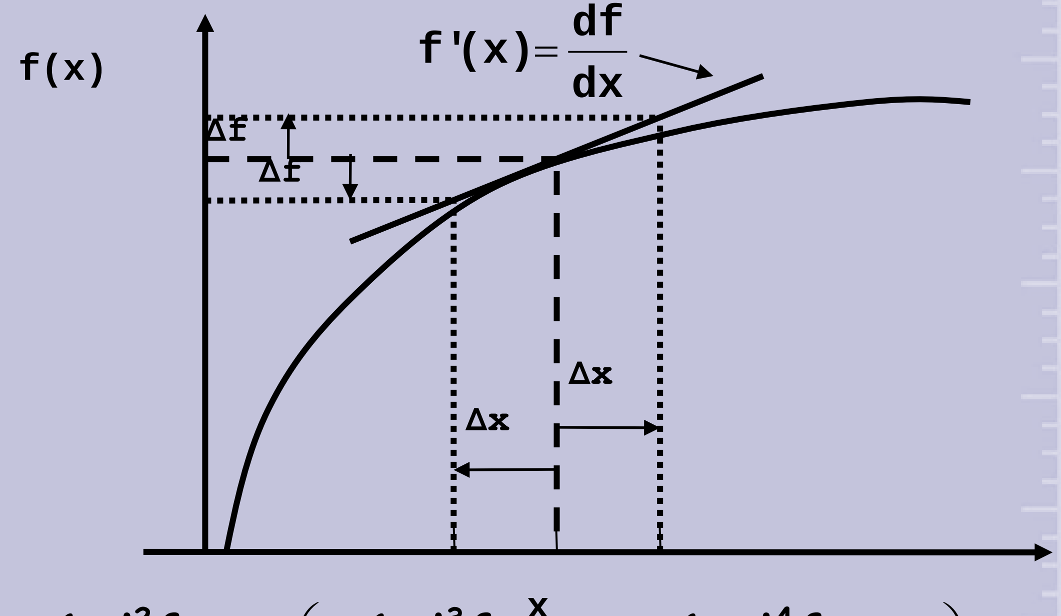
α=1 gives (approximately) a 68% confidence interval

α=2 gives 95% CL (and at least 75% from Bienaymé-Chebyshev)

Error propagation

Measure : $x \pm \Delta x$

Compute : $f(x) \rightarrow \Delta f$?



Assuming **small errors**,

using Taylor expansion :

$$f(x+\Delta x) = f(x) + \frac{df}{dx}\Delta x + \frac{1}{2}\frac{d^2f}{dx^2}\Delta x^2 \left(+ \frac{1}{6}\frac{d^3f}{dx^3}\Delta x^3 + \frac{1}{24}\frac{d^4f}{dx^4}\Delta x^4 \right)$$

$$f(x-\Delta x) = f(x) - \frac{df}{dx}\Delta x + \frac{1}{2}\frac{d^2f}{dx^2}\Delta x^2 \left(- \frac{1}{6}\frac{d^3f}{dx^3}\Delta x^3 + \frac{1}{24}\frac{d^4f}{dx^4}\Delta x^4 \right)$$

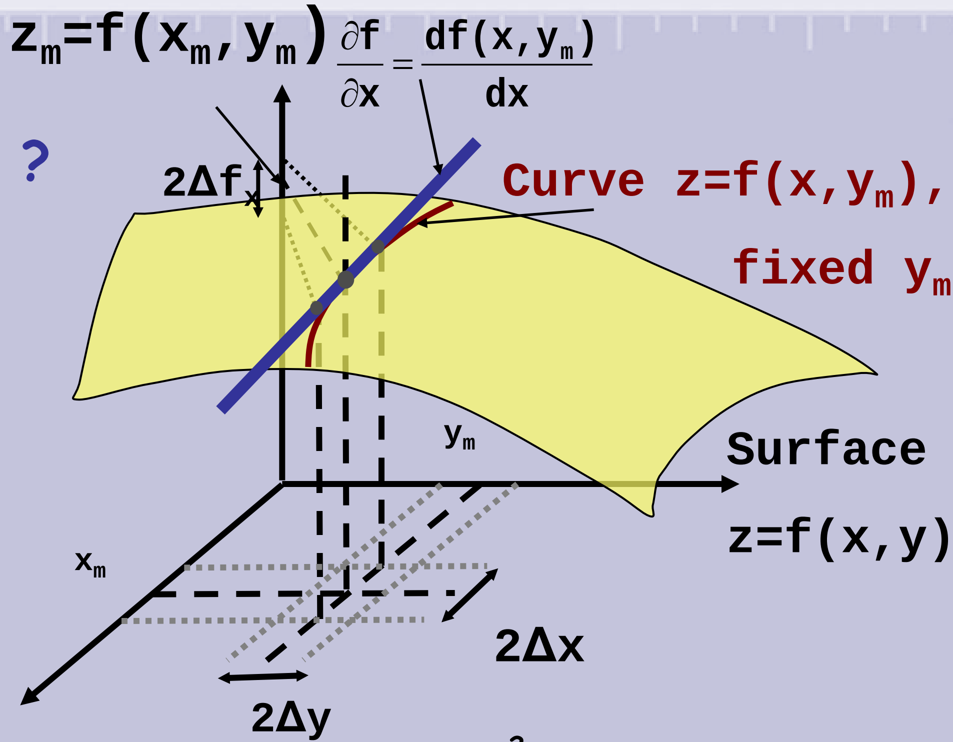
$$\Rightarrow \Delta f = \frac{1}{2}\left| f(x+\Delta x) - f(x-\Delta x) \right| = \frac{df}{dx}\Delta x \left(+ \frac{1}{6}\frac{d^3f}{dx^3}\Delta x^3 \right)$$

Error propagation

Measure : $x \pm \Delta x,\ y \pm \Delta y, \ldots$

Compute : $f(x,y,\ldots) \rightarrow \Delta f$?

Idea : treat the effect of each variable as separate **error sources**



$$\Delta_x f = \left|\frac{\partial f}{\partial x}\right| \Delta x \ , \ \Delta_y f = \left|\frac{\partial f}{\partial y}\right| \Delta y$$

Then

$$\Delta f^2 = \Delta_x f^2 + \Delta_x f^2 + \rho_{xy} \Delta_x f \Delta_y f = \left(\frac{\partial f}{\partial x}\Delta x\right)^2 + \left(\frac{\partial f}{\partial y}\Delta y\right)^2 + \rho_{xy} \frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\Delta x \Delta y$$

$$\Delta f^2 = \sum_i \left(\frac{\partial f}{\partial x_i}\Delta x_i\right)^2 + \sum_{i,j<i} \rho_{x_i x_j} \left|\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\right| \Delta x_i \Delta x_j$$

uncorrelated

$$\Delta f^2 = \sum_i \left(\frac{\partial f}{\partial x_i}\Delta x_i\right)^2$$

correlated

$$\Delta f = \left|\frac{\partial f}{\partial x}\right|\Delta x + \left|\frac{\partial f}{\partial y}\right|\Delta y$$

anticorrelated

$$\Delta f = \left|\left|\frac{\partial f}{\partial x}\right|\Delta x - \left|\frac{\partial f}{\partial y}\right|\Delta y\right|$$