THEORY #3

Statistical Method & Techniques for Data Analysis

Erasmus+ EU programme

Alexis Pompili (*)
University of Bari Aldo Moro

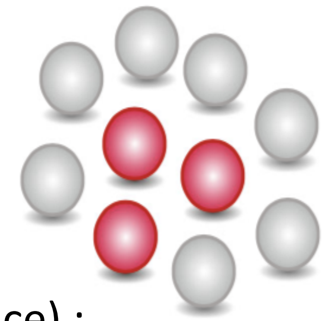
Erasmus+ Teaching Mobility Program / 2025 May 6-13 @ NKUA (Physics Dpt.)

(*) alexis.pompili@ba.infn.it
(or alexis.pompili@cern.ch)

PART 3A - CORE

BERNOUILLI TRIALS & BINOMIAL DISTRIBUTION

Bernoulli distribution - I



- Let's consider a basket containing a number of balls, each having one of 2 possible colors (red and white).

Assume we know the number R of red balls in the basket and the number W of white balls (R=3, W=7).

The probability to randomly extract a red ball in basket is, according to the expression of the classical (Laplace) :

$$p(R) = R/(R+W) = 3/10$$

A variable $\boldsymbol{k}$ equal to the number of red balls in one extraction (called **Bernouilli trial**) can assume only the values 0 or 1 and is called **Bernouilli variable** :

$$\begin{cases} p(k=1) = p \\ p(k=0) = q = 1-p \end{cases} \quad , \quad p(k=0) + p(k=1) = p + q = p + 1 - p = 1$$

Thus the (discrete) distribution probability of $\boldsymbol{k}$ (**Bernouilli distribution**) is given by: $P(k) = p^k \cdot q^{1-k} = p^k \cdot (1-p)^{1-k}$

Bernoulli trial - II

- Expectation value and variance of the discrete random variable k are easily calculated:

$$\mu_k \equiv E[k] = \frac{\sum_{k=0}^{1} kP(k)}{\sum_{k=0}^{1} P(k)} = \frac{\cancel{0 \cdot q} + 1 \cdot p}{q+p} = \frac{p}{q+p} = \frac{p}{1-p+p} = p$$

$$\sigma_k^2 \equiv V[k] = E\left[k^2\right] - (E[k])^2 = \frac{\sum_{k=0}^{1} k^2 P(k)}{\sum_{k=0}^{1} P(k)} - \mu_k^2 = \frac{\cancel{0^2 \cdot q} + 12 \cdot p}{q+p} - p^2 = p - p^2 = (1-p)p = qp$$

- The next step is to consider what happens when considering a sequence of N independent Bernouilli trials, and ask ourselves which is the probability to have, for instance, n heads when lunching the coin N times.

  This **extension** is actually called **binomial distribution** and it's discussed in the next slides.

  Example with $N = 2$: $p(m=0) + p(m=1) + p(m=2) = q^2 + (pq + qp) + p^2 = (p+q)^2 = (p+1-p)^2 = 1$

- A ***binomial process*** consists of a given number N of **independent** Bernoulli trials, each with probability p.
  This could be implemented, for instance, by randomly extracting a ball from a basket containing a fraction p of red balls; after each extraction, the extracted ball is placed again in the basket (so that the subsequent extraction is not biased by the outcome of the previous one) and then the extraction is repeated, for a total of N extractions.

The **binomial probability distribution** provides the probability of k successes when N independent Bernouilli trials are carried out, each trial characterized by a probability p of success (and a probability $1-p$ of failure).

Since the trials are **independent** …the probability of **any** sequence of successes (S) and failures (F) in some defined order is equal to the product of the single probabilities !
Ex.: with $N=5$, the probability associated to the sequence SSFSF is given by $pp(1-p)p(1-p)=p^3(1-p)^2$.

In general the probab. of a particular sequence with k successes & $N-k$ failures is given by: $p^k \cdot q^{N-k} = p^k \cdot (1-p)^{N-k}$

-

Binomial distribution - I

- A ***binomial process*** consists of a given number $\boldsymbol{N}$ of **independent** Bernoulli trials, each with probability $\boldsymbol{p}$.
  This could be implemented, for instance, by randomly extracting a ball from a basket containing a fraction $\boldsymbol{p}$ of red balls; after each extraction, the extracted ball is placed again in the basket (so that the subsequent extraction is not biased by the outcome of the previous one) and then the extraction is repeated, for a total of $\boldsymbol{N}$ extractions.

The **binomial probability distribution** provides the probability of $\boldsymbol{k}$ successes when $\boldsymbol{N}$ independent Bernouilli trials are carried out, each trial characterized by a probability $\boldsymbol{p}$ of success (and a probability $\boldsymbol{1-p}$ of failure).

Since the trials are **independent** …the probability of **any** sequence of successes (S) and failures (F) in some defined order is equal to the product of the single probabilities !

Ex.: with $N = 5$, the probability associated to the sequence SSFSF is given by $\boldsymbol{pp(1-p)p(1-p) = p^3(1-p)^2}$.

In general the probab. of a particular sequence with $\boldsymbol{k}$ successes & $\boldsymbol{N-k}$ failures is given by: $p^k \cdot q^{N-k} = p^k \cdot (1-p)^{N-k}$

- Since **the order of successes and failures is not important** we need to consider, in the evaluation of the probability of $\boldsymbol{k}$ success in $\boldsymbol{N}$ trials, **all** the different sequences, i.e. all the possible combinations of $\boldsymbol{k}$ success and $\boldsymbol{N-k}$ failures in $\boldsymbol{N}$ trials.

  For this purpose we need to remember that if we have $\boldsymbol{N}$ distinguishible objects and we want to group them in group of $\boldsymbol{k}$ each time, we have that the number of combinations when the order is not important, is given by:

$$C_{N,k} = \binom{N}{k} = \frac{N!}{k!(N-k)!}$$

- To conclude we can state that:

The total probability to have k successes in N trials is given by the binomial probability distribution

$$\boldsymbol{B(\underbrace{k}_{\text{discrete random variable}};\underbrace{N,p}_{\text{distribution parameters}}) = \binom{N}{k} p^k (1-p)^{N-k}} \qquad \text{with} \quad \binom{N}{k} = \frac{N!}{k!\,(N-k)!}$$

- It can be verified that the probability distribution $B(k;N,p)$is properly normalized:

$$\sum_{k=0}^{N} B(k;N,p) = \sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k} = [\ldots] = (p+q)^N = (p+1-p)^N = 1$$

- The binomial probability is symmetric when $p = q = 1/2$: $B\left(k;N,\frac{1}{2}\right) = \binom{N}{k}\frac{1}{2^k}\frac{1}{2^{N-k}} = \binom{N}{k}\frac{1}{2^N}$ is an **odd function**

Binomial distribution - III

- The most important properties of the binomial distribution are :

Expectation value $\mu = E[k] = Np$

Variance $\sigma^2 = V[k] = Np(1-p) = Npq$

Formal demonstrations are given in the additional material.

- A typical application of a binomial distribution is any type of efficiency:
  - Detection efficiency (of a detector)
  - Reconstruction efficiency (of a reconstruction software)
  - Selection efficiency (of an analysis to extract a signal from the background)

N. The binomial probability distribution is shown in Fig. 2.1 and Fig. 2.2 for various values of p and N.

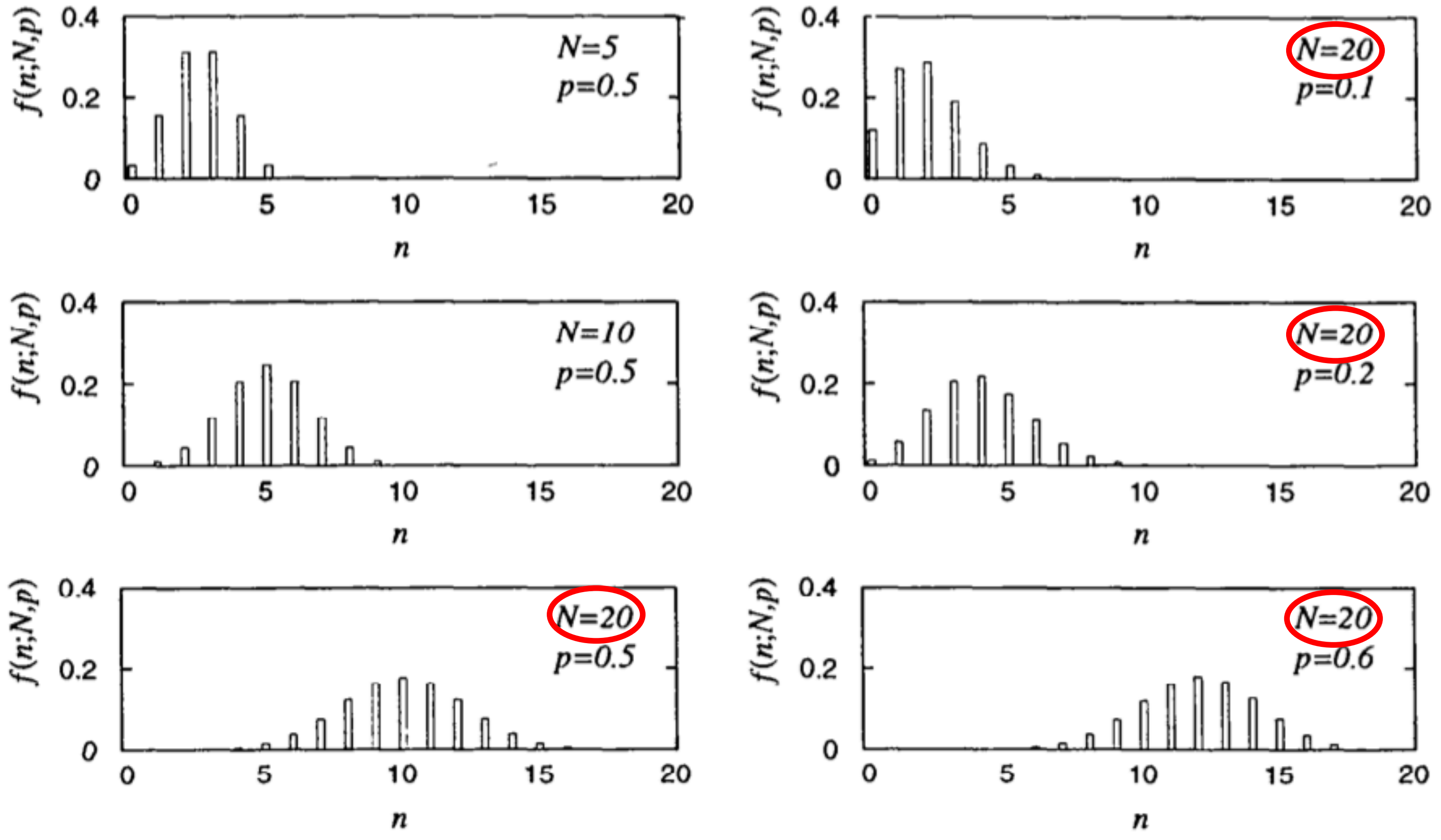


Fig. 2.1 The binomial distribution for $p = 0.5$ and various values of N.

Fig. 2.2 The binomial distribution for $N = 20$ and various values of p.

Binomial distribution for the *relative* number of successes

- Often we are interested to the random variable $\frac{k}{N}$ that is the relative number of successes for N trials .

Thus we interested to $\frac{k}{N}$ instead of k. However considering that N is simply a constant, we can write:

Expectation value :

$$E\left[\frac{k}{N}\right] = \frac{1}{N}E[k] = \frac{pN}{N} = p$$

Note: the parameter p is the expectation value of the fraction of successes,
which of course is intuitively expected since p represents the probability of a success in 1 trial.

Variance :

$$V\left[\frac{k}{N}\right] = \frac{1}{N^2}V[k] = \frac{Npq}{N^2} = \frac{pq}{N} \equiv \frac{p(1-p)}{N}$$

Note: the variance V tends to reduce and become null in the limit of infinite number of trials
(since p and q are finite).

Example : estimate of the efficiency of a particle detector - I

- Particles' detectors are examples of such devices: they produce a signal when a particle interacts with them, but they may fail to do this in a certain fraction of times.

The distribution of the number of positive signals $\boldsymbol{n}$, if $\boldsymbol{N}$ processes of interest occured, ...

is given by a **binomial distribution** with parameter $p = \varepsilon$.

A typical example is represented by the estimate of the **efficiency ε of a device**.

A way to estimate the efficiency consists in performing a large number $\boldsymbol{N}$ of sampling of the process of interest,

counting the number of times the device gives a positive signal (i.e. it has been efficient).

In a typical **test beam for a particle detector** the data acquisition time should be sufficiently long in order to get a large number of particle crossing the detector.

Example : estimate of the efficiency of a particle detector - II

- Let us assume that the result of a real experiment of N particles crossing the detector gives a measured value of n equal to $\hat{n}$; an estimate of the true efficiency ε is given by ...

$$\hat{\varepsilon} = \frac{\hat{n}}{N}$$

The uncertainty on the estimate of the true efficiency is given by ... $\sigma_{\hat{\varepsilon}} = \sqrt{\mathrm{V}[\hat{\varepsilon}]} \equiv \sqrt{\mathrm{V}\left[\frac{\hat{n}}{N}\right]} = \sqrt{\frac{V[\hat{n}]}{N^2}} = \sqrt{\frac{N\varepsilon(1-\varepsilon)}{N^2}}$

But this is not very useful since the true efficiency ε is unknown !

Anyway, if N is sufficiently large we can assume - with good approximation - that $\hat{\varepsilon}$ will be very close to the true efficiency ε (as a consequence of the law of large numbers), and thus by replacing ε with $\hat{\varepsilon}$ we get the following approximated expression for the uncertainty:

$$\sigma_{\hat{\varepsilon}} \cong \sqrt{\frac{\hat{\varepsilon}\,(1-\hat{\varepsilon})}{N}}$$

Note that the above formula leads to an error in the extreme cases [when $\hat{\varepsilon} = 0$ & $\hat{\varepsilon} = 1$ i.e. for $\hat{n} = 0$ & $\hat{n} = N$].
A solution to the problem of determining the correct confidence interval for a binomial distribution is due to Copper & Pearson (correspoding to the Neyman inversion of a confidence belt)[**see hands-on exercise**]

Similar example : selection efficiency - I

- We get a typical application of the binomial pdf whenever we want to discriminate among a signal & its backgrounds by using the information on a generic variable *x* & requiring that an event is selected if satisfies the selection criterium $\mathbf{x} > \mathbf{X}_{\mathbf{cut}}$

The selection efficiency $\boldsymbol{\varepsilon}$ can be defined as the fraction of the events (in the limit of infinite events analysed) that satisfies the selection criterium.

Since an event either satisfies the criterium or fails to be selected, the number of selected events, N_{sel} , is distributed according to a binomial pdf:

$$\boldsymbol{B\left(\frac{N_{sel}}{N_{tot}}; N_{tot}, p\right)}$$

where the probability $\mathbf{p}$ represents the fraction of successes after infinite trials, namely the selection efficiency $\boldsymbol{\varepsilon}$ by definition, and $\boldsymbol{N_{tot}}$ is the number of events/trials (supposed very large).

Similarly to what already discussed one gets for the expectation value: $$E\left[\frac{N_{sel}}{N_{tot}}\right] = \frac{1}{N_{tot}} \cdot E\left[N_{sel}\right] = \frac{1}{N_{tot}} \cdot \varepsilon N_{tot} = \varepsilon$$

… and gets for the variance: $$V\left[\frac{N_{sel}}{N_{tot}}\right] = \frac{1}{N_{tot}^2} \cdot V\left[N_{sel}\right] = \frac{1}{N_{tot}^2} \cdot \varepsilon(1-\varepsilon) N_{tot} = \frac{\varepsilon(1-\varepsilon)}{N_{tot}}$$

- Very often the selection efficiency is estimated, by using simulated data (Monte Carlo data), as the ratio between selected simulated events (n) and total simulated events (N); however, nobody has infinite simulated statistics and N can still be enough large but never infinite.

  The estimation of the efficiency in this real case (N not infinite) will be given by: $\hat{\varepsilon} = \dfrac{n}{N}$

  ...and the uncertainty on its estimation by: $\sigma_{\hat{\varepsilon}} = \sqrt{\dfrac{\hat{\varepsilon}(1-\hat{\varepsilon})}{N}}$

Note that the uncertainty on the efficiency estimation decreases for increasing number of simulated events:

the more Monte Carlo you produce the more precise efficiency estimation you get!

Tipically **the statistical error in a Monte Carlo estimation becomes a systematic error in real data analysis**

(as we discussed earlier)!

Multinomial Distribution

- It is possible to generalize the binomial distribution to the case where there are more than 2 possible results.
  Let's suppose we can have $\boldsymbol{m}$ different results with associated probabilities $p_i \geq 0\ \forall i\ (i = 1, \dots, m)$.
  Let's consider now N trials and denote with $\boldsymbol{k_i}$ the number of trials having the result labelled by index i.

  We can now **extend the binomal p.d.f. at this case** writing down the new p.d.f.:

$$\mathbf{M}(k_1, \dots, k_m; N, p_1, \dots, p_m) = \frac{N!}{k_1! \cdot \dots \cdot k_m!} \cdot p_1^{k_1} \cdot \dots \cdot p_m^{k_m}$$

Multinomial pdf

… where the normalization conditions are the following: $\sum_{i=1}^{m} p_i = 1$, $\sum_{i=1}^{m} k_i = N$ (this is the additional condition when N is fixed)

We can now rewrite this **multinomial distribution** in a more compact form:

$$\boxed{\mathbf{M}(\vec{k}; N, \vec{p}\,) = N! \cdot \prod_{i=1}^{m} \frac{p_i^{k_i}}{k_i!}}$$

An example of application of this distribution is the histogram of $\boldsymbol{m}$ bins with probability $p_i \geq 0\ \ (i = 1, \dots, m)$ that an event enters the bin i. For N events (resulting from N trials), the probability that the number of events in each of the $\boldsymbol{m}$ bins will be given by $\boldsymbol{k_i}$ for each bin $i = 1, \dots, m$ is provided by the multinomial distribution!

The expectation values and the variances of the multinomial variables $\boldsymbol{k_i}$ are obtained by considering that for the bin i a generic event either "drops" inside it (with probability p_i) or does not (with probability $1-p_i$) so that the $\boldsymbol{k_i}$ variable is **binomal** taken singularly and therefore it holds:

$$\boldsymbol{\mu_i \equiv \mathrm{E}[k_i] = Np_i} \qquad \boldsymbol{\sigma_i^2 \equiv \mathrm{V}[k_i] = Np_i(1-p_i) = Np_i \sum_{j\neq i} p_j}$$

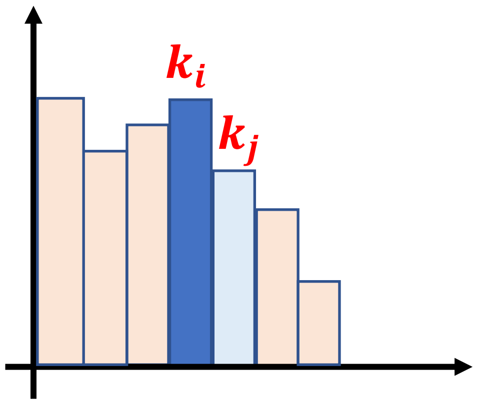
$$\boxed{\sum_{i=1}^{m} p_i = 1}$$

Moreover, since is N fixed the condition $\sum_{i=1}^{m} k_i = N$ implies that the $\boldsymbol{k_i}$ numbers are not independent among each other!

Thus we need to introduce the covariance between any possible couple $(\boldsymbol{k_i}, \boldsymbol{k_j})$ with $\boldsymbol{j \neq i}$.

It can be demonstrated that :

$$\mathrm{cov}(\boldsymbol{k_i}, \boldsymbol{k_j}) = \boldsymbol{-Np_ip_j} \qquad \text{for each pair } (\boldsymbol{k_i}, \boldsymbol{k_j})$$



The sign "-" indicates (as intuitively we would expect) that

… the number of events in two possible generic bins $(\boldsymbol{k_i}, \boldsymbol{k_j})$ are negatively correlated!

POISSON DISTRIBUTION

Poisson distribution - I

- The binomial distribution has an interesting limit when N is very large and p is very small …
… so that the product $\nu = Np$ (called the ***rate*** parameter) is finite.

We start from the binomial random variable k and we re-parametrize its distribution using ν instead of p :

$$B(k; N, \nu) = \binom{N}{k} \left(\frac{\nu}{N}\right)^k \left(1 - \frac{\nu}{N}\right)^{N-k}$$

that can be explicitly rewritten as

$$B(k; N, \nu) = \frac{N!}{k!(N-k)!} \left(\frac{\nu}{N}\right)^k \left(1 - \frac{\nu}{N}\right)^{N-k}$$

… and conveniently rewritten as:

$$B(k; N, \nu) = \left(\frac{\nu^k}{k!}\right) \frac{N(N-1)\dots(N-k+1)}{N^k} \left(1 - \frac{\nu}{N}\right)^N \left(1 - \frac{\nu}{N}\right)^{-k}$$

depends on N

We get the limit for $N \to \infty$ of the 3 terms in the red ellipse:

$$\lim_{N\to\infty} \frac{\overbrace{\boxed{N(N-1)\dots(N-k+1)}}^{k \text{ terms}}}{N^k} \cong 1 \quad , \quad \lim_{N\to\infty} \left(1 - \frac{\nu}{N}\right)^N = \lim_{N\to\infty} e^{N\ln(1-\nu/N)} = e^{-\nu} \quad , \quad \lim_{N\to\infty} \left(1 - \frac{\nu}{N}\right)^{-k} = 1 \text{ because } \nu/N \to 0$$

Thus, in this limit, the binomial distribution becomes the Poisson distribution: $\mathrm{P}(k; \nu) = \dfrac{\nu^k e^{-\nu}}{k!}$

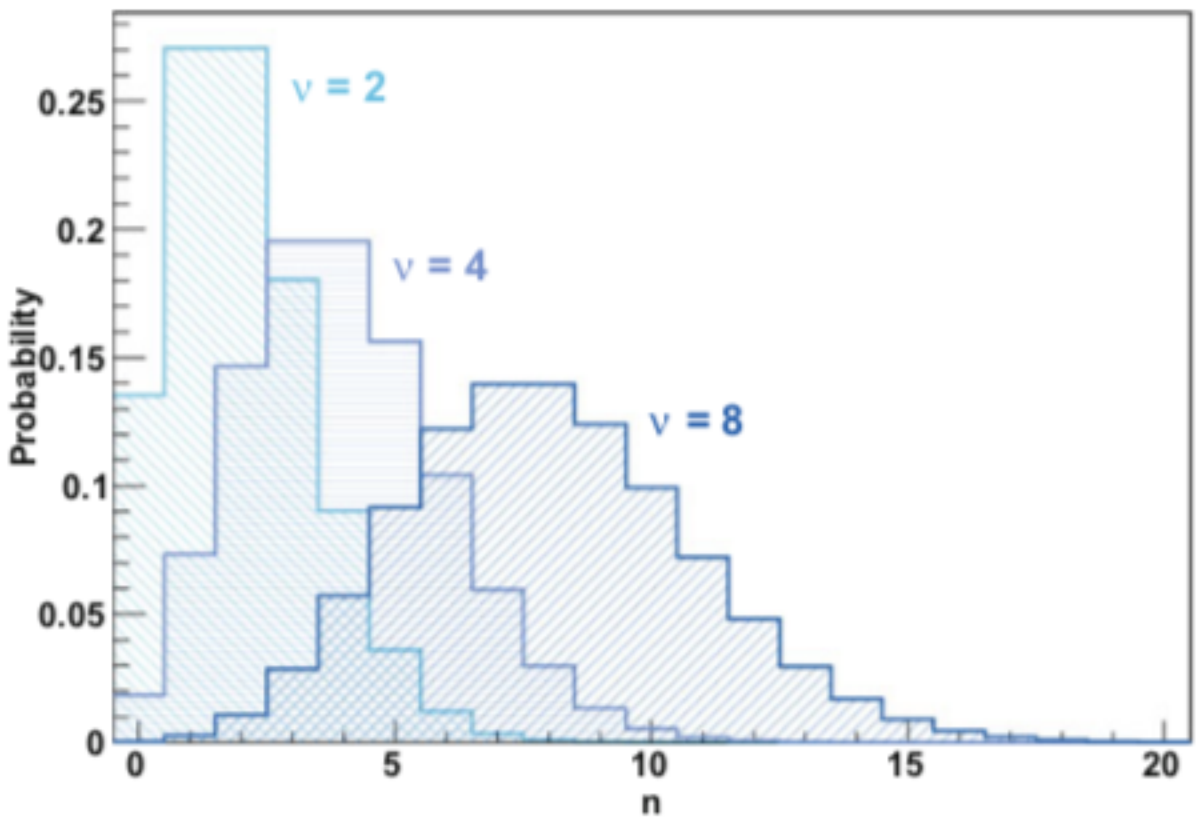
A non-negative integer r.v. k is called Poissonian r.v. if it is distributed, for a given value of the parameter ν, according to it!

Poisson distribution

A non-negative integer random variable n is called *Poissonian random variable* if it is distributed, for a given value of the parameter ν, according to the distribution in Eq. (2.70). A Poissonian random variable has expected value and variance both equal to ν. This descends directly from the limit of a binomial distribution. Figure 2.5 shows examples of Poisson distributions for different values of the parameter ν.

$$\mu_n \equiv \mathrm{E}[n] = \nu$$

$$\sigma_n^2 \equiv \mathrm{V}[n] = \nu$$

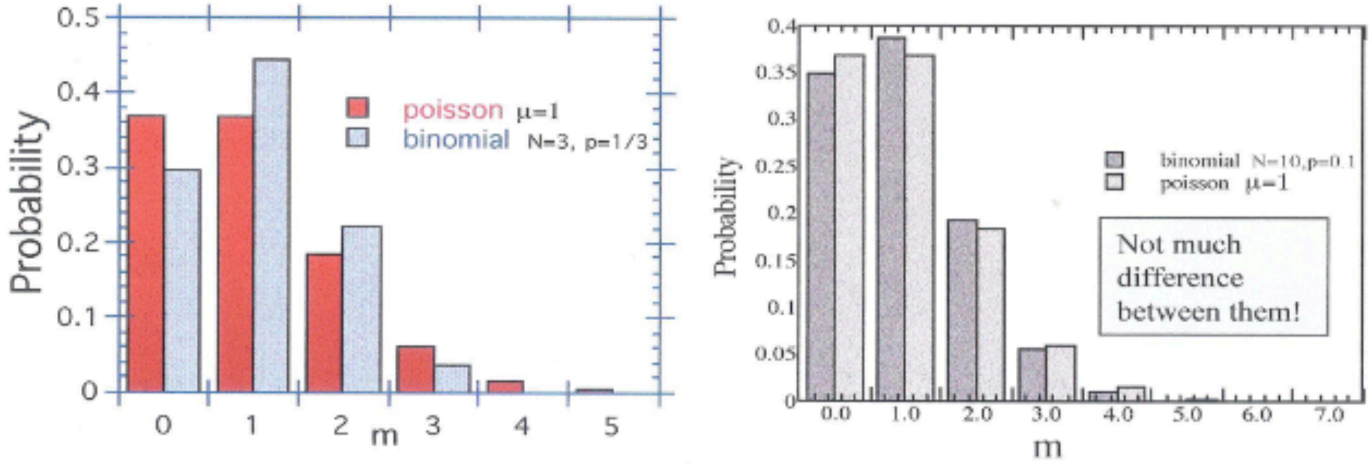


Poisson distributions with different value of the rate parameter ν

Poisson distributions have several interesting properties, some of which are listed below:

- A binomial distribution with a number of extractions N and probability $p \ll 1$ can be approximated with a Poisson distribution with average $\nu = p\,N$, as discussed above.

Comparison of Binomial and Poisson distributions with mean $\mu = 1$



For large N: Binomial distribution looks like a Poisson of the same mean

- For large ν, a Poisson distribution can be approximated with a Gaussian distribution (see Sect. 3.25) having average ν and standard deviation $\sqrt{\nu}$. See Fig. 2.6 for a visual comparison.

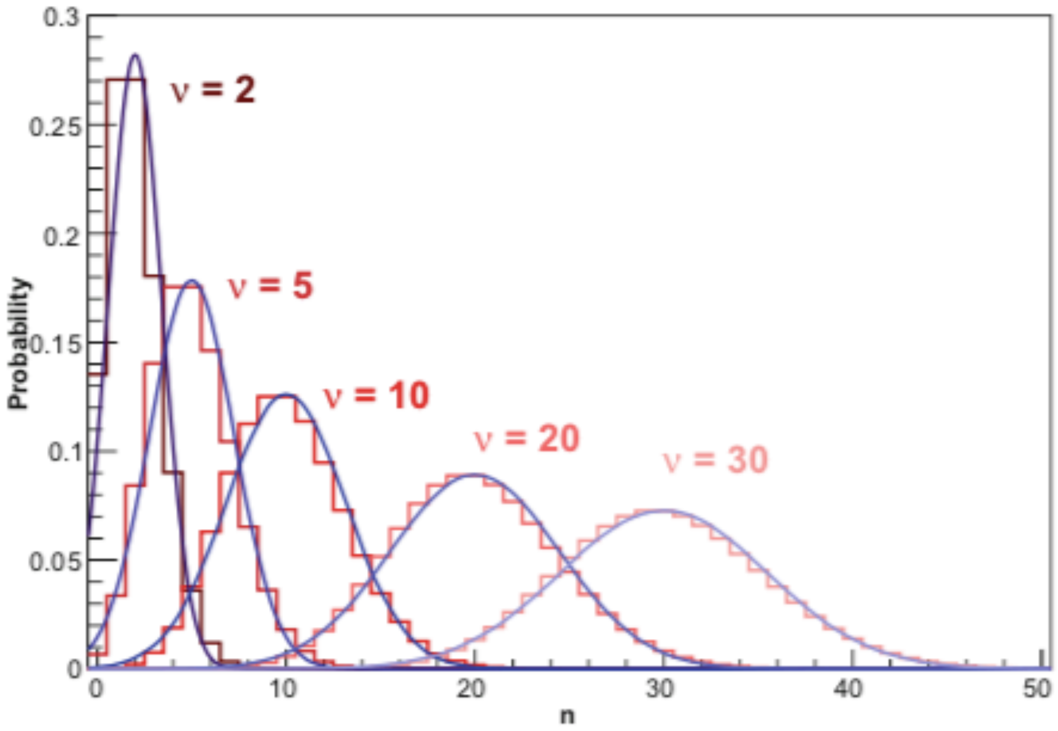


Fig. 2.6 Poisson distributions with different value of the parameter ν compared with Gaussian distributions with $\mu = \nu$ and $\sigma = \sqrt{\nu}$

Histogram as application of a Poisson distribution

- A set of r.v.s that we "meet everyday" in Physics is represented by a so-called **histogram** (introduced when discussing the multinomial distribution)

Let's consider a r.v. $\boldsymbol{x}$, characterized by an underlying distribution $\boldsymbol{f(x)}$, of which we carry out a sampling of $\boldsymbol{N}$ values (i.e. we perform $\boldsymbol{N}$ measurements) $\{x_j\}_{j=1,\dots,N}$; it's possible to subdivide these values into $\boldsymbol{N_B}$ intervals.

We denote the interval labelled with index $\boldsymbol{i}$ as the bin $\boldsymbol{I_i}$ with $\boldsymbol{i} = 1, \dots, \boldsymbol{N_B}$. In the bin $\boldsymbol{I_i}$ "fall" $\boldsymbol{N_i}$ values of $\boldsymbol{x}$.

Of course $\boldsymbol{N_i}$ are so that $\sum_{i=1}^{N_B} N_i = N$ holds. The relevant aspect here is that these numbers $\boldsymbol{N_i}$ are r.v.s themselves !

In the hypothesis to have:

1) a probability enough small that a measurement "falls" in a certain bin rather than in any other ($\boldsymbol{N_B}$ enough large)
2) the number of registered events, populating the histogram, rather large ($\boldsymbol{N}$ large allows $\boldsymbol{N_B}$ enough large)

... these r.v.s $\boldsymbol{N_i}$ are distributed according to the poissonian distribution (for each bin):

$$\mathrm{P}(N_i;\mu_i) = \frac{\mu_i^{N_i} e^{-\mu_i}}{N_i!}$$

... where $\boldsymbol{\mu_i}$ is the expectation value associated to the bin $\boldsymbol{i}$:

$$\mu_i = \int_{I_i} f(\xi)d\xi$$

Application: histogram with the number of entries ($\boldsymbol{N}$) not fixed (otherwise it's a multinomial distribution across the bins)

Poisson distribution : applicability - I

The poissonian distribution is applicable in the following circumstances:

a) the events are independent (and are in a number sufficiently large)

b) the "event rate" is constant

We call stocastic processes all those processes for which the events are independent & happen at a constant rate

and a typical example are the radiative decays (provided the observation time is well shorter than the half-life of the source!)

When Poissonian is not applicable? 3 examples:

1) counting experiment of radiative decays from a **relatively too small radioactive source**

2) counting experiment of radiative decays for a **relative too long time (i.e.comparable) w.r.t. the radioactive source lifetime**

⇨ in these 2 cases **the event rate is not constant** (diminishes with time)

3) counting the interactions produced by a beam consisting of a relative small # of particles impinging a relatively thick target

⇨ in this case **the event rate is not constant** because it diminishes with the depth of penetration of the beam in the target

Poisson distribution : applicability - II

Another limitation to the stocastic nature of counting a particle flux by means of a detector/counter (a counting equipment) is introduced by the so-called ***dead time*** **of a detector/counter**.

Any counter is not capable of counting more than 10^6 particles per second, because they are characterized by a ***dead time*** not below the μsec.

If the particles flux is relatively not too high … the probability that a 2nd particle crosses the counter during the dead time triggered from the crossing of a 1st particle is negligible and - consequently - the detection of the 2nd particle is independent from that of the 1st.

Instead, if the flux is too high ($> 10^6 p/sec$) the detections of the particles are no more independent …
…. and Poisson statisitcs does not apply.

Poisson distribution : applicability - III

Add….

Gaussian Distribution

A *Gaussian distribution*, or *normal distribution*, is defined by the following PDF, where μ and σ are fixed parameters:

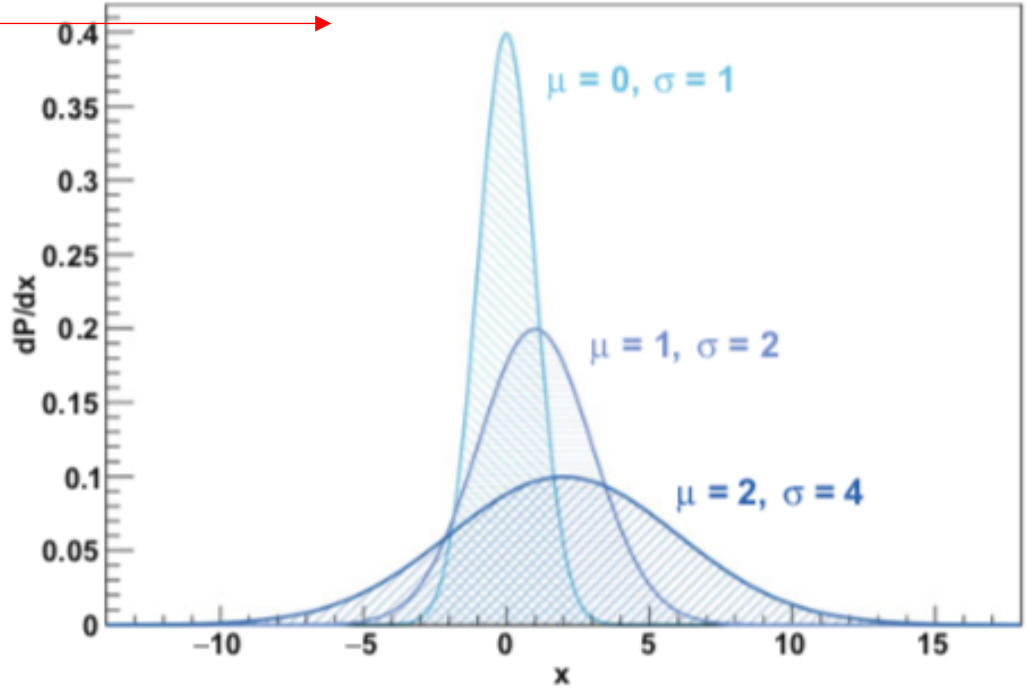
$$\boxed{p(x;\ \mu,\ \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\ .}$$

symmetric w.r.t. μ

A random variable distributed according to normal distribution is called *Gaussian random variable*, or *normal random variable*. The function has a single maximum at $x = \mu$ and two inflection points at $x = \mu \pm \sigma$. Examples of Gaussian distributions are shown in Fig for different values of μ and σ.

The average value and standard deviation of a normal variable are μ and σ respectively. The *full width at half maximum* (FWHM) of a Gaussian distribution is equal to $2\sigma\sqrt{2\log 2} \simeq 2.3548\,\sigma$.

$$\begin{cases} \mu_x \equiv \mathrm{E}[x] = \mu \\ \sigma_x^2 \equiv \mathrm{V}[n] = \sigma^2 \end{cases}$$



Gaussian p.d.f - II

- For $\mu = 0$ and $\sigma = 1$, a normal distribution is called *standard normal* or *standard Gaussian* distribution. A *standard normal random variable* is often indicated with z, and its distribution is:

$$\boxed{\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}}$$

If we have a normal random variable x with average μ and standard deviation σ, a standard normal variable can be built with the following transformation:

$$z = \frac{x - \mu}{\sigma}$$

- The cumulative distribution of a standard normal distribution is: $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z'^2/2}\, dz' = \frac{1}{2}\left[\mathrm{erf}\left(\frac{z}{\sqrt{2}}\right) + 1\right]$

The probability for a Gaussian distribution corresponding to the symmetric interval around μ $[\mu - Z\sigma,\ \mu + Z\sigma]$, frequently used in many applications, can be computed as:

$$P(Z\sigma) = \frac{1}{\sqrt{2\pi}} \int_{-Z}^{Z} e^{-x^2/2}\, dx = \Phi(Z) - \Phi(-Z) = \mathrm{erf}\left(\frac{Z}{\sqrt{2}}\right) . \qquad (3.50)$$

The most frequently used values are the ones corresponding to 1σ, 2σ, and 3σ $(Z = 1, 2, 3)$ and have probabilities of 68.27%, 95.45% and 99.73%, respectively.

error function

$$\mathrm{erf}(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x'^2}{2}}\, dx'$$

(reminder)

(evaluated numerically)

Gaussian p.d.f - III

- To verify the normalization:

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\frac{\xi^2}{\sigma^2}}\, d\xi = \frac{1}{\sqrt{2\pi}\cdot\sigma}\int_{-\infty}^{+\infty} e^{-\frac{\xi^2}{2\sigma^2}}\, d\xi \quad \left[\int_{-\infty}^{+\infty} e^{-\frac{x^2}{c}}\,dx = \sqrt{\pi c}\right] \quad \cancel{\frac{1}{\sqrt{2\pi}\cdot\sigma}}\cdot\cancel{\sqrt{\pi\, 2\sigma^2}} = 1$$

- The importance of the Gaussian distribution resides in the central limit theorem (see Sect. 3.20), which allows to approximate to Gaussian distributions may realistic distributions resulting from the superposition of more random effects.

Chi-squared Distribution

A χ^2 *random variable* with k *degrees of freedom* is the sum of the squares of k independent standard normal variables $z_1, \cdots, z_k$, defined above in Sect. 3.8:

$$\chi^2 = \sum_{j=1}^{k} z_i^2 . \tag{3.51}$$

If we have k independent normal random variables $x_1, \cdots, x_k$, each with mean μ_j and standard deviation σ_j, a χ^2 variable can be built by transforming each x_j in a standard normal variable using the transformation in Eq. (3.49):

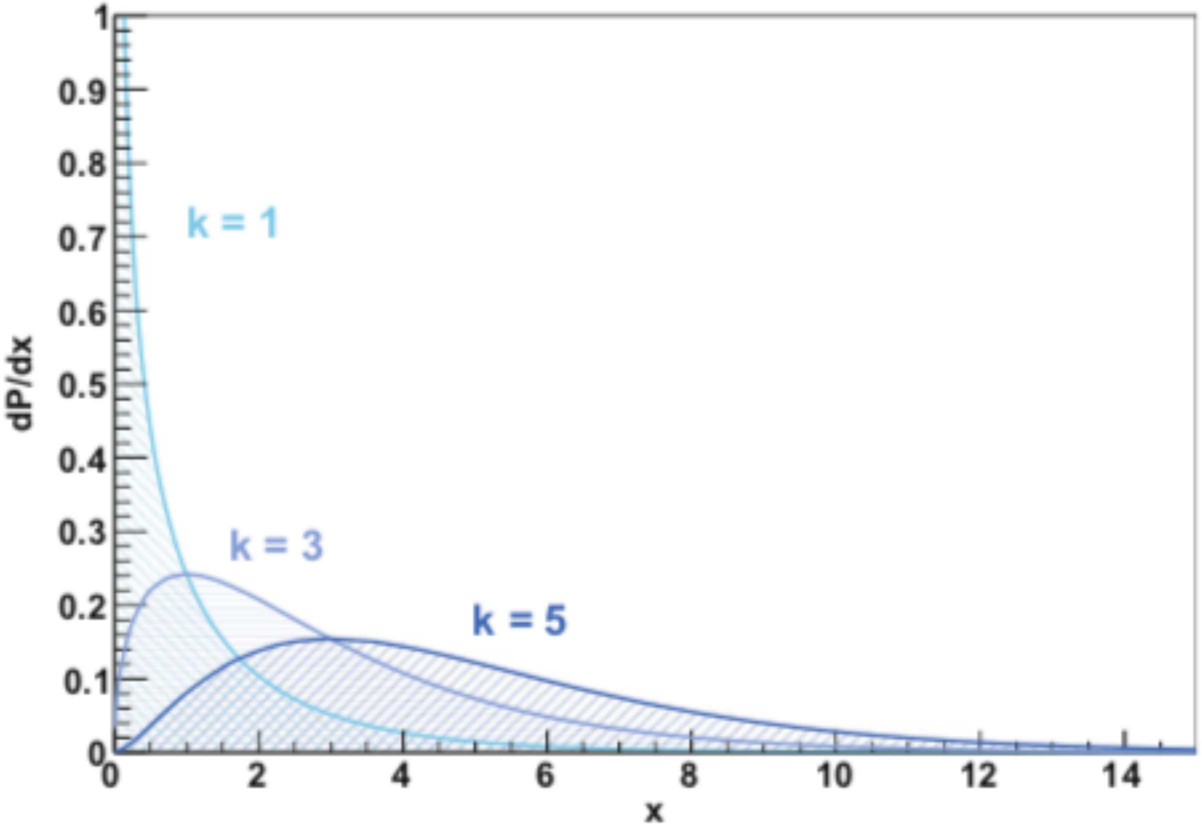
$$\chi^2 = \sum_{j=1}^{k} \frac{(x_j - \mu_j)^2}{\sigma_j^2} . \tag{3.52}$$

The distribution of a χ^2 variable with k degrees of freedom is given by:

$$\boxed{p(\chi^2; k) = \frac{2^{-k/2}}{\Gamma(k/2)} \chi^{k-2} e^{-\chi^2/2} .} \tag{3.53}$$

Γ is the so-called *gamma function* and is the analytical extension of the factorial.[1] Sometimes, the notation χ_k^2 is used to explicitate the number of degrees of freedom in the subscript k. χ^2 distributions are shown in Fig. 3.3 for different number of degrees of freedom k.

Chi-squared p.d.f - II



χ^2 distributions with different numbers of degrees of freedom k

$$\begin{cases} \mu_x \equiv \mathrm{E}[x] = k \\ \sigma_x^2 \equiv \mathrm{V}[n] = 2k \end{cases}$$

The expected value of a χ^2 variable is equal to the number of degrees of freedom k and the variance is equal to $2k$. The cumulative χ^2 distribution is given by:

$$P(\chi^2 < \nu) = \int_0^{\nu} p(\chi^2;\, k)\,\mathrm{d}\,\chi^2 = P\left(\frac{k}{2},\, \frac{\nu}{2}\right), \tag{3.54}$$

where $P(a,\, k) = \gamma(a,\, k)/\Gamma(a)$ is the so-called *regularized Gamma function*. It is related to the Poisson distribution by the following formula:

$$\sum_{j=0}^{k-1} \frac{e^{-\nu}\nu^j}{j!} = 1 - P(k,\, \nu), \tag{3.55}$$

so that the following relation holds:

$$\sum_{j=0}^{k-1} \frac{e^{-\nu}\nu^j}{j!} = \int_{2\nu}^{+\infty} p(\chi^2;\, 2k)\,\mathrm{d}\,\chi^2. \tag{3.56}$$

Chi-squared p.d.f - III

- Typical applications of the χ^2 distribution are $\boxed{\text{goodness-of-fit tests}}$ (see Sect. 6.17), where the cumulative χ^2 distribution is used.

We will use it in the hands-on exercises

Application: minimization tasks can be carried out using a chi-squared but this can be applied only to binned distributions (i.e.only for histograms). Historically it has been superseeded in minimization tasks by the likelihood method (which can be used also for unbinned distirbutions) [as discussed in the following heory part]

Central Limit Theorem - I

There are more or less complex/rigorous ways to enunciate this theorem.

I choose here one of them in a form that usually can also be demonstrated (but in this context I will omit the demonstration)

Given n **independent** random variables $\{x_i\}_{i=1,\dots,n}$, distributed according to the p.d.f.s $\{f_i(x_i)\}_{i=1,\dots,n}$, with expectation values $\{\mu_i\}_{i=1,\dots,n}$ and variances $\{V_i = \sigma_i^2\}_{i=1,\dots,n}$,

(generally can be a *weighted* linear combination)

… the random variable sum of the $\{x_i\}_{i=1,\dots,n}$, denoted as $S = \sum_{i=1}^{n} x_i$, will be characterized by:

an expectation value $\mu_S \equiv E[S] = \sum_{i=1}^{n} \mu_i$ and variance $\sigma_S^2 \equiv V[S] = \sum_{i=1}^{n} \sigma_i^2$

and the associated p.d.f. $f[S]$ will *tend to* the **normal p.d.f.** for $n \to \infty$: $\lim_{n\to\infty} f[S] = \mathbb{N}\left(S; \sum_{i=1}^{n} \mu_i , \sum_{i=1}^{n} \sigma_i^2\right)$

"converges in law"

… provided that the expectation value and variance **exist**,
the single variances are **finite** and are of the **<u>same</u> order of magnitude**: $\frac{\sigma_{x_i}}{\sigma_{x_j}} = O(1) \quad \forall (i,j)$ pair

Note: the demonstration would be enough affordable assuming the simplifying case for which $f_i = f \; \forall i$

- What does mean "convergence in law" (also called "convergence in distribution")?

A set of variables $\{x_n\}$ converges in law to a variable x if,
denoted the correspoding c.d.f.s as $F_{x_n}(X)$ and $F_x(X)$, it holds: $\lim_{n\to\infty} F_{x_n}(X) = F_x(X) \quad \forall X$ in which F_x is continuous

(Note: it is weaker than "Convergence in probability")

- What is the meaning of the theorem?

The theorem holds for sum of **r.v.s**, both discrete and continuous, having different p.d.f.s! However it's essential that …
… they are **independent** and that, *each one taken singularly has **weak** impact on the overall final result*.

The theorem states that :
a r.v. tends to be distributed according to a Gaussian p.d.f. IF it's a linear superposition of a large number independent r.v.s that taken singularly have a weak influence on the final result.

In other words…
the theorem assigns to the Gaussian p.d.f. the role of universal function to which tend all the p.d.f.s of r.v.s of systems "of high statistical equilbrium".

PART 3B - IN-DEPTH SLIDES

- How to demonstrate that: $\mu = E[k] = Np$

1ª) $E[k] = \sum_{k=0}^{N} k\, B(k;N,p) = \sum_{k=0}^{N} k \binom{N}{k} p^k (1-p)^{N-k} = \sum_{k=0}^{N} k \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k} =$

(IL TERMINE CON $k=0$ È NULLO!)

$$= \sum_{k=1}^{N} k \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k} = \sum_{k=1}^{N} \frac{N(N-1)!\,\cancel{k}}{\cancel{k}(k-1)!\,(N-k_{-1+1})!}\, p\, p^{k-1} (1-p)^{N-k-1+1} =$$

(DEFINISCO ED INTRODUCO $N' = N-1$, $k' = k-1$, SI HA: $\sum_{k=1}^{N} \rightarrow \sum_{k'=0}^{N'}$)

$$= Np \sum_{k'=0}^{N'} \frac{N'!}{k'!(N'-k')!} p^{k'} (1-p)^{N'-k'}$$

$$= Np \underbrace{\sum_{k'=0}^{N'} B(k';N',p)}_{=1\ [\text{NORMALIZZAZIONE di } B(k';N',p)]} = Np$$

- How to demonstrate that: $\sigma^2 = V[k] = Np(1-p) = Npq$

2°) RICORDIAMOCI CHE PER LA R.V. x (CONTINUA) : $V[x] = E[(x-\mu)^2] = E[x^2] - (E[x])^2$

ABBIAMO APPENA CALCOLATO $E[k]$ E QUINDI NON "RIMANE" CHE CALCOLARE $E[k^2]$:

$$E[k^2] = \sum_{k=0}^{N} k^2 B(k;N,p) = \sum_{k=0}^{N} k^2 \binom{N}{k} p^k (1-p)^{N-k} = \sum_{k=0}^{N} k^2 \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k}$$

(IL TERMINE CON $k=0$ È NULLO!)

$$= \sum_{k=1}^{N} k^2 \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k} = \sum_{k=1}^{N} \frac{N(N-1)!\,k^{\cancel{2}}}{\cancel{k}(k-1)!(N-k-1+1)!}\, p\, p^{k-1} (1-p)^{N-k-1+1}$$

$$= Np \sum_{k=1}^{N} \frac{(N-1)!\,(k-1+1)}{(k-1)!\,(N-1-k+1)!} p^{k-1} (1-p)^{N-1-k+1} =$$

DEFINISCO ED INTRODUCO $N' = N-1$, $k' = k-1$, SI HA: $\sum_{k=1}^{N} \longrightarrow \sum_{k'=0}^{N'}$

$$= Np \sum_{k'=0}^{N'} \frac{N'!\,(k'+1)}{k'!\,(N'-k')!} p^{k'} (1-p)^{N'-k'}$$

$$E[k^2] = Np \sum_{k'=0}^{N'} (k'+1) \frac{N'!}{k'!(N'-k')!} p^{k'} (1-p)^{N'-k'} =$$

$$= Np \Bigg[\underbrace{\sum_{k'=0}^{N'} k' \frac{N'!}{k'!(N'-k')!} p^{k'} (1-p)^{N'-k'}}_{\sum_{k'=0}^{N'} k' B(k';N',p) = E[k'] = N'p} + \underbrace{\sum_{k'=0}^{N'} \frac{N'!}{k'!(N'-k')!} p^{k'} (1-p)^{N'-k'}}_{=1} \Bigg]$$

[NORMALIZZAZIONE di $B(k';N',p)$]

$$= Np\,(N'p+1) \underset{(N'=N-1)}{=\!=} Np[(N-1)p+1] = Np\,(Np-p+1)$$

in conclusione:

$$V[k] = E[k^2] - (E[k])^2 = Np\,(Np-p+1) - (Np)^2$$

$$= \cancel{(Np)^2} + Np(1-p) - \cancel{(Np)^2} = Npq$$

c.v.d.