Chapter 2

Statistics and the Treatment of Experimental Data

statistics plays an essential part in all the sciences.

Many of the process involved with detection of particles are statistical in nature.

- For the experimentalist, it is also a design and planning tool.

before performing any measurement, one must consider the tolerances required of the apparatus, the measuring times involved, etc., as a function of the desired precision on the result. Such an analysis is essential in order to determine its feasibility in material, time and cost.

- The understanding and interpretation of all experimental data depend on statistical and probabilistic concepts.

Characterization of data

A collection of N *independent measurements* of the same physical quantity:

$$x_1, x_2, x_3, \ldots, x_i, \ldots, x_N$$

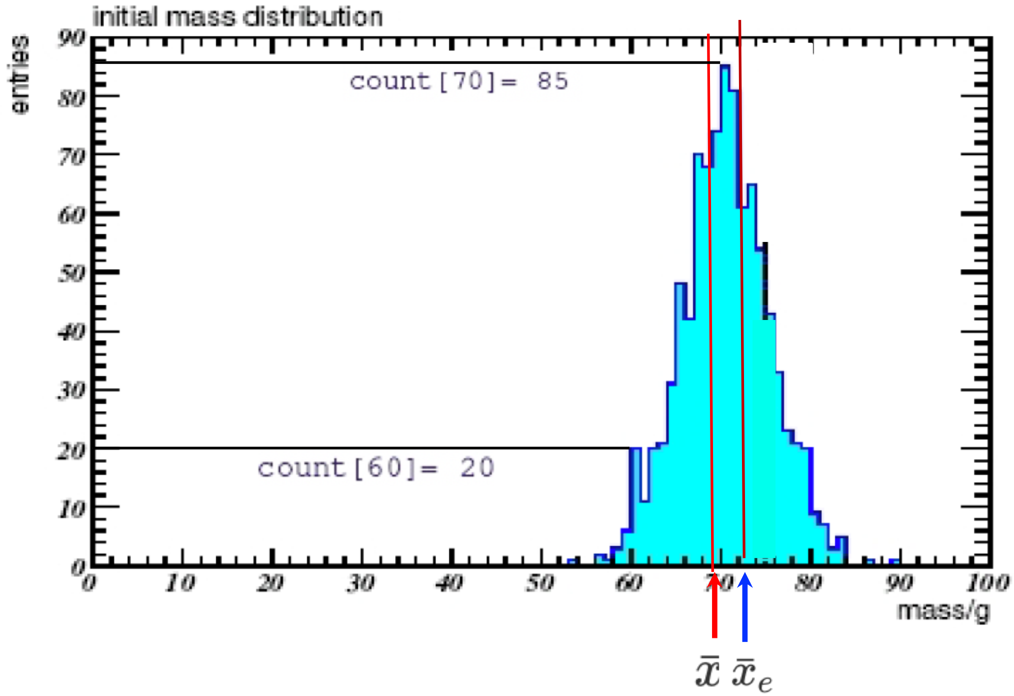
```
73 79 72 62 67 60 60 67 78 68 66 75 76 73 75 64 70 69 73 59 70 73 64 72 64 69
69 71 69 71 77 69 72 71 67 72 63 66 68 76 71 76 68 71 63 65 65 66 73 73 73 67
70 65 71 69 78 67 65 69 71 71 72 73 72 69 66 66 70 60 72 62 53 65 74 65 68 69
67 75 64 76 72 76 78 67 67 67 69 79 71 67 71 68 71 65 66 65 78 76 71 70 67 65
67 64 73 67 74 79 74 71 73 67 66 76 68 74 76 65 77 67 71 67 71 77 63 66 70 62
68 74 67 67 67 77 65 68 79 72 71 77 68 70 73 67 81 70 74 71 79 62 67 63 68 76
73 81 76 73 68 72 76 61 69 73 71 80 68 70 62 76 58 68 68 64 68 78 69 65 70 70
64 75 73 72 60 86 68 68 64 60 68 71 70 75 70 67 69 67 73 65 66 71 70 70 73 66
72 71 71 64 76 75 72 72 71 72 72 71 75 68 73 70 64 76 72 75 79 70 64 70 67 70
75 70 83 69 61 70 66 69 71 72 70 76 73 62 71 60 73 74 70 68 68 70 78 71 69 71
73 73 75 65 71 67 60 70 77 71 74 64 74 73 60 77 73 70 69 66 70 78 69 75 66 71
75 75 74 69 74 70 75 77 75 66 72 68 72 61 75 65 69 68 65 73 82 67 75 67 80 71
79 72 71 68 73 70 67 75 74 69 63 63 72 70 73 63 70 70 59 78 76 66 72 79 65 71
76 72 69 69 73 70 77 73 83 66 68 67 69 73 76 65 71 70 71 65 78 71 67 70 72 75
67 79 72 64 62 79 68 70 61 65 68 71 73 60 60 68 71 74 75 69 73 70 68 ...
```

Data Reduction - **Counter**

```
count[50]=  0     count[60]= 20     count[70]= 85     count[80]=  9
count[51]=  0     count[61]= 11     count[71]= 81     count[81]=  7
count[52]=  0     count[62]= 20     count[72]= 61     count[82]=  3
count[53]=  0     count[63]= 21     count[73]= 65     count[83]=  5
count[54]=  0     count[64]= 31     count[74]= 54     count[84]=  0
count[55]=  0     count[65]= 48     count[75]= 43     count[85]=  0
count[56]=  2     count[66]= 42     count[76]= 33     count[86]=  1
count[57]=  1     count[67]= 70     count[77]= 23     count[87]=  0
count[58]=  3     count[68]= 68     count[78]= 21     count[88]=  0
count[59]=  6     count[69]= 74     count[79]= 20     count[89]=  1
```

Histogram

- define bins for the possible values of a variable
- plot the number of entries in each bin



Experimental mean: $\bar{x}_e = \sum_{i=1}^{N} x_i / N$

$$P(x_i) = \frac{N_{x_i}}{N}$$

$$\bar{x}_e = \sum_i x_i P(x_i)$$

True mean value: $\bar{x}$

Residual：残差

$$\epsilon_i \equiv x_i - \bar{x}$$

Sample variance: 样本方差

An index of the degree of the internal scatter in data

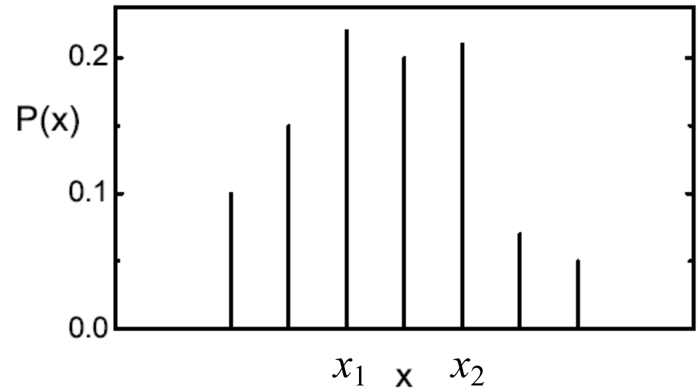
$$s^2 \equiv \overline{\epsilon^2} \doteq \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 \qquad \bar{x} \to \bar{x}_e \qquad s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x}_e)^2$$

Characteristics of Probability Distributions

The probability of finding x between certain limits, $P(x_1 \leq x \leq x_2)$

Discrete distributions

$$P(x_1 \leq x \leq x_2) = \sum_{i=1}^{2} P(x_i), \qquad \sum_{i} P(x_i) = 1$$



Continuous distributions

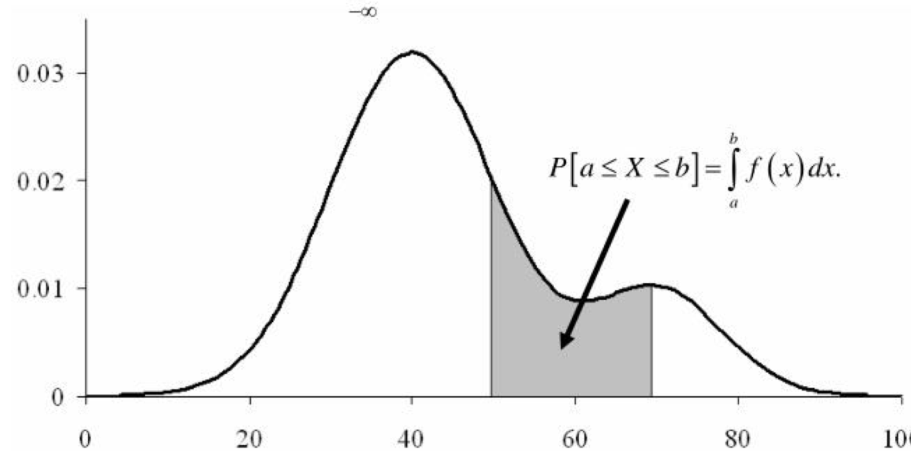
The probability to measure a value x

in the interval [x, x+dx] is given by

probability density function f(x) (p.d.f)

Probability density function, $f(x)$

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$



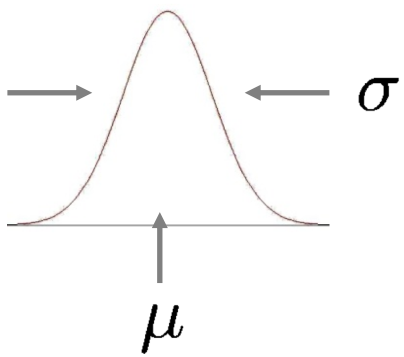
Expectation(mean) values

Discrete $$\mu = E[x] = \sum x_i P(x_i)$$

Continuous $$\mu = E[x] = \int x f(x) dx$$

Variance: 方差

$$\sigma^2 = E[(x-\mu)^2] = \int (x-\mu)^2 f(x) dx$$



Standard deviation: 标准偏差

$$\sigma = \sqrt{\sigma^2}$$

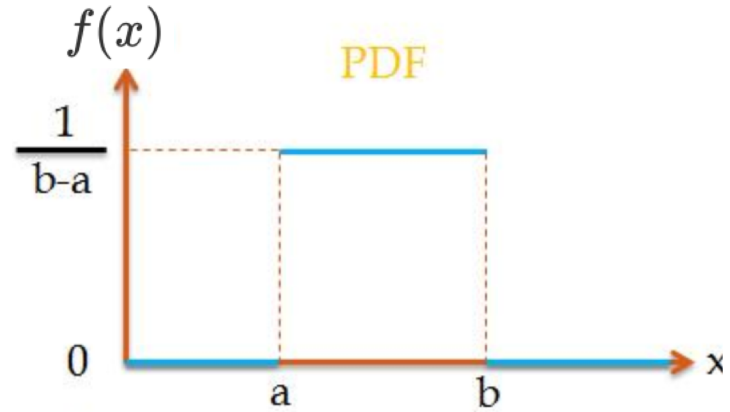
Measure for dispersion/spread of a distribution is given by variance around a single mean μ

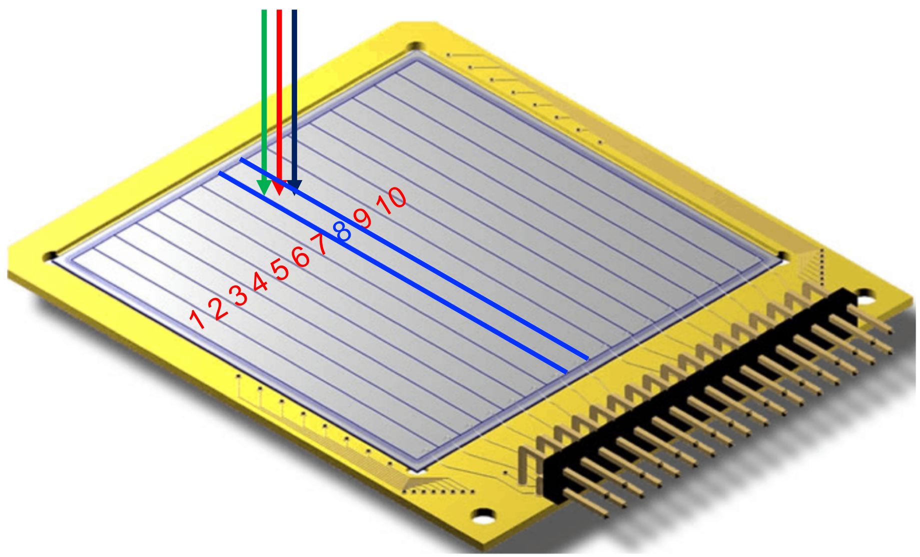
Report result: $\mu \pm \sigma$

Example: Uniform distribution: U(a,b)

$$\mu = \int x f(x) dx = \frac{b+a}{2}$$

$$\sigma^2 = \int (x-\mu)^2 f(x) dx = \frac{(b-a)^2}{12}$$





Silicon Strip Detectors:

Resolution of one strip cluster: $\frac{\Delta x}{\sqrt{12}}$

$$8 \pm 1/\sqrt{12}$$

Some common probability distributions

Consider *N* independent experiments (Bernoulli trials):

1. The experiment has two possible outcomes, sucess(s) and faliure(f).
2. The probability that any given observation results in an outcome of type s or f is constant, independent of the number of observations.

Trial	Definition of a success	Probability of a success
Toss of a coin	"Heads"	1/2
Toss of a die	"A four"	1/6
Observation of a radioactive nucleus for a time "t"	It decays	$1 - e^{-\lambda t}$
Observation of a detector of efficiency E placed near a radioactive nucleus for a time "t"	A count	$E(1 - e^{-\lambda t})$

$$N(t) = N_0 e^{-\lambda t}$$

$$\Delta N(t) = N_0 (1 - e^{-\lambda t})$$

$$p = \frac{\Delta N(t)}{N} = 1 - e^{-\lambda t}$$

The Binomial Distribution 二项式分布

probability of success on any given trial is p.

Probability of a specific outcome (in order), e.g. '*ssfsf*' is

$$pp(1-p)p(1-p) = p^3(1-p)^{5-3}$$

Probability of **r** successes in **n** trails (in order), is

$$p^r(1-p)^{n-r}$$

regardless of the order,

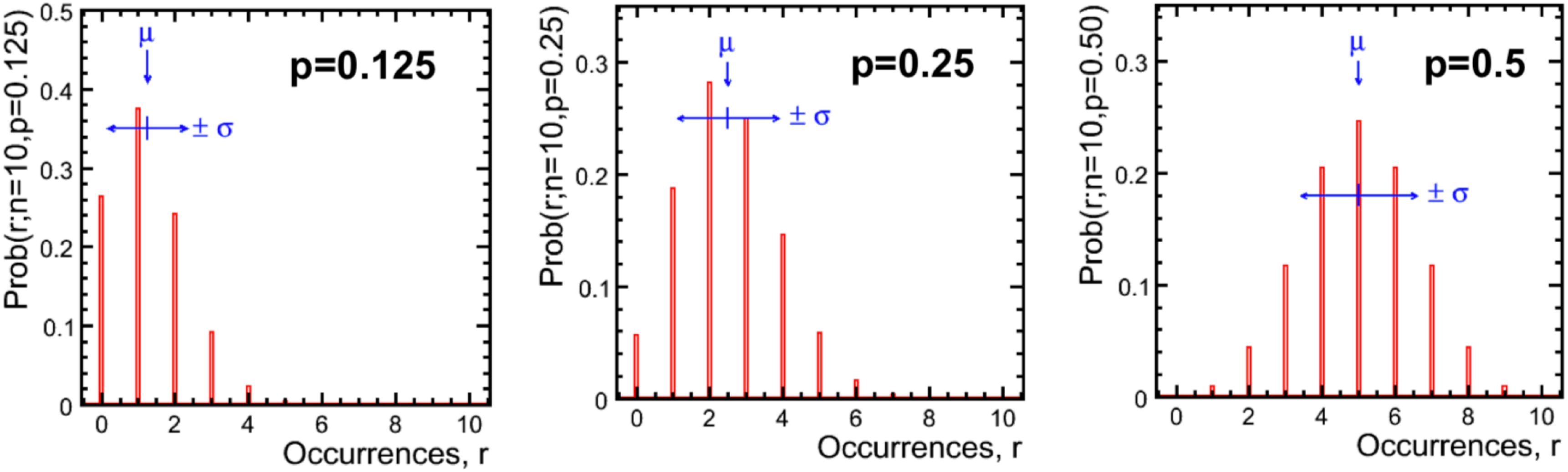
$$P(r;p,n) = \frac{n!}{r!(n-r)!}p^r(1-p)^{n-r}$$

the expectation value and variance

$$\mu = np$$

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

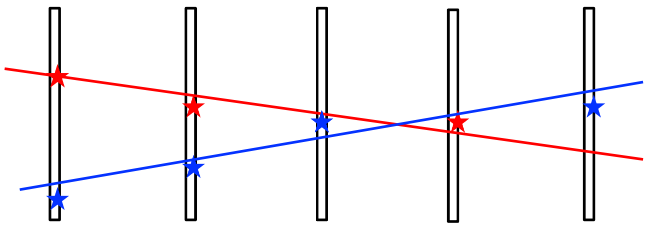
e.g. n=10



- The distribution is not symmetric.
- The peak or maximum of the distribution does not correspond to the mean.

- Binomial distribution is the most general model and is widely applicable to all constant-p processes.

Example – Detector Design



- Particle goes through a detector layer.
  - "success" measuring signal (p), - efficiency
  - "failure" no meas. (1-p)

efficiency $= \frac{N_{det}}{N_{inc}}$

- How many layer do I need to have high overall track finding efficiency ? (3 points without magnetic field, with B field at least 4)

- Assume efficiency: p = 95%
  - 3 layers: P(3;3,0.95) = 0.857
  - 4 layers: P(3;4,0.95) + P(4;4,0.95) = 0.986
  - 5 layers: P(3;5,0.95) + P(4;5,0.95) + P(5;5,0.95) = 0.999

- Redundancy is very important when building detectors, assume always worst case

The Poisson Distribution 泊松分布

The Poisson distribution occurs as the limiting the binomial distribution when the probability $p \rightarrow 0$ (unknown) and the number of trials $N \rightarrow \infty$, such that the constant average rate $\mu = Np$, remains finite.

The probability of observing r events in this limit then reduces to

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

The standard deviation

$$\sigma = [\sum_r (r-\mu)^2 P(r)]^{1/2} = \sqrt{\mu}$$

This is the origin of the formula $n \pm \sqrt{n}$

The Poisson distribution depends on only one parameter: μ, so that knowledge of N and p is not always necessary.

Examples:

- number of a specific type of events in a particle-particle scattering, when the total number of events is large and this specific process is very rare: e.g. Higgs decay

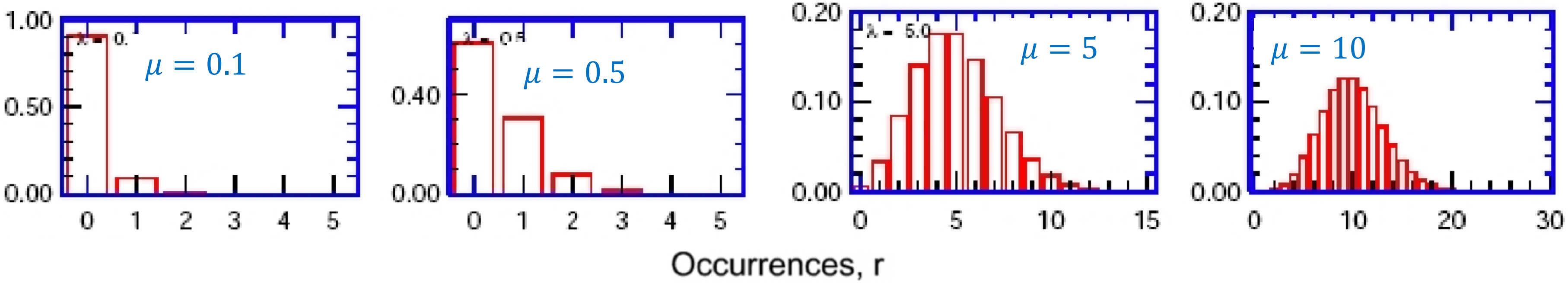
An example of radioactive decay.

Decay of 25 mg of an element, with the lifetime of 10^{12} years .

N= 10^{20} atoms (very large), $T_{1/2}$=5x10^{19} seconds.

The probability of a given nucleus to decay , $\lambda=\ln 2/T_{1/2}$= 2x10^{-20}/sec (very small).

Np = 2 (finite!)

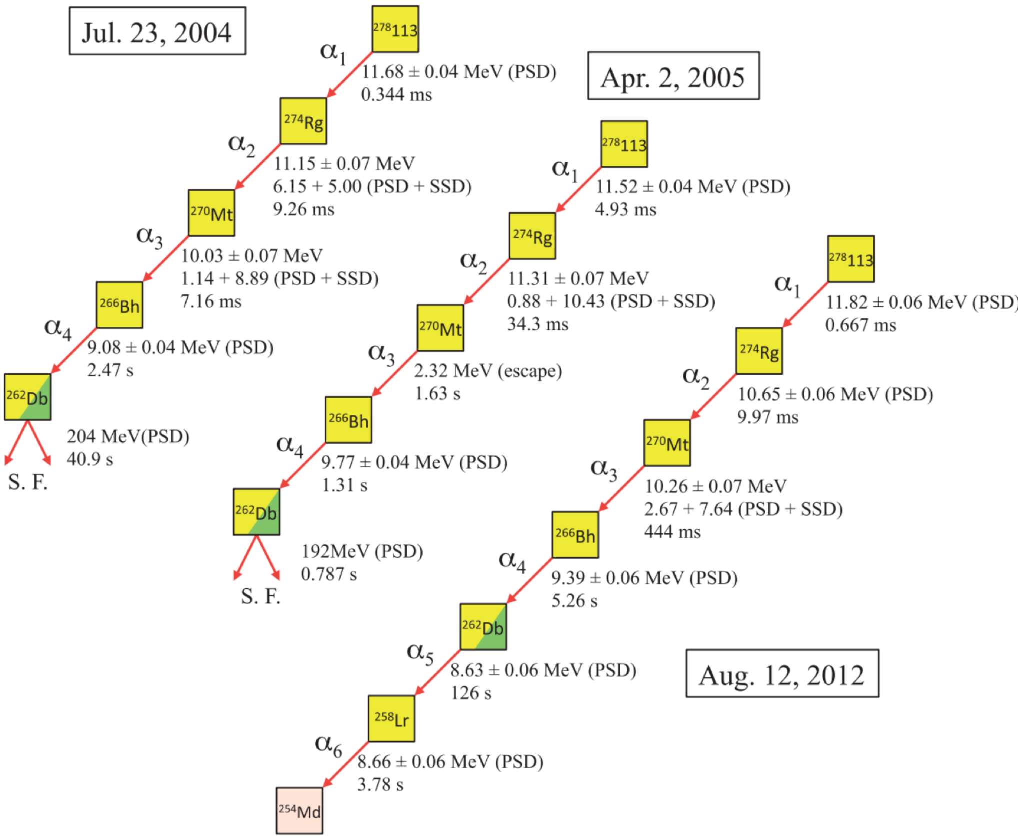


- The distribution is not symmetric.
- The peak or maximum of the distribution does not correspond to the mean.

μ< 1: most probable result is 0

As μ becomes large, the distribution becomes more and more symmetric and approaches a Gaussian form(μ≧20).

Example: synthesis of superheavy element: Z=113



JPSP 81 (2012) 103201

Table I. Summary of beamtime used.

Beamtime		Irradiation time (days)	Beam dose/sum ($\times 10^{19}$)	Number of observed events
year	month/day			
2003	9/5–12/29	57.9	1.24/1.24	0
2004	7/8–8/2	21.9	0.51/1.75	1
2005	1/20–1/23	3.0	0.07/1.82	0
2005	3/20–4/22	27.1	0.71/2.53	1
2005	5/19–5/21	2.0	0.05/2.58	0
2005	8/7–8/25	16.1	0.45/3.03	0
2005	9/7–10/20	39.0	1.17/4.20	0
2005	11/25–12/15	19.5	0.63/4.83	0
2006	3/14–5/15	54.2	1.37/6.20	0
2008	1/9–3/31	70.9	2.28/8.48	0
2010	9/7–10/18	30.9	0.52/9.00	0
2011	1/22–5/22	89.8	2.01/11.01	0
2011	12/2–12/19	14.4	0.33/11.34	0
2012	1/15–2/9	25.0	0.56/11.90	0
2012	3/13–4/17	33.7	0.79/12.69	0
2012	6/12–7/2	15.7	0.25/12.94	0
2012	7/14–8/18	32.0	0.57/13.51	1
Total	Total	553	13.51	3

Very low probability p(σ~50 fb), large N
-> very small mean Np. Np=3/553= 0.0054/day
Very large probability of no-events.

Poisson random variable (x)	0
Average rate of success	0.0054
Poisson Probability: P(X = 0)	0.99461
Cumulative Probability: P(X > 0)	0.00539

Example: Neutrinos from Supernovae

- Irvine-Michigan-Brookhaven experiment looking for neutrinos 23/2/1987, about which time supernova 1987a exploded.
- Table of distributions of number of events observed in 10s time intervals:

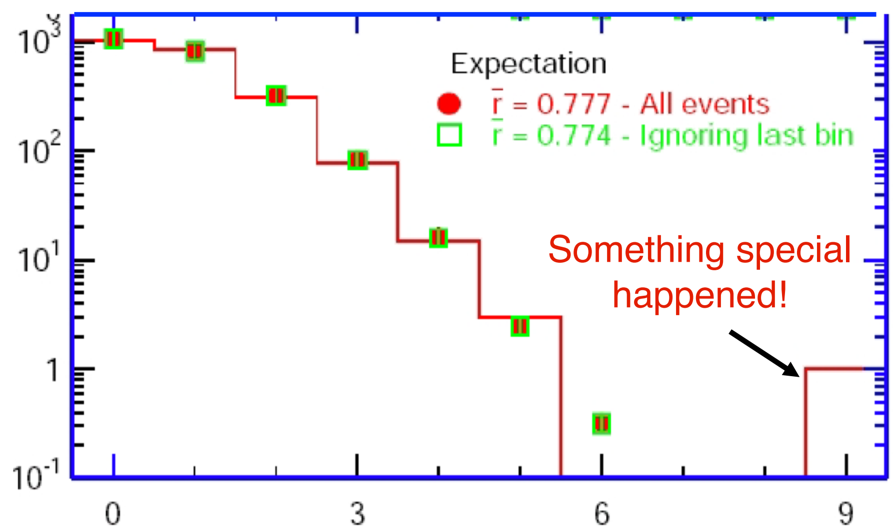
No. of events	0	1	2	3	4	5	6	7	8	9
No. of intervals	1042	860	307	78	15	3	0	0	0	1
Prediction	1064	823	318	82	16	2	0.3	0.03	0.003	0.0003

What is the mean number?

μ = 0.777;

excluding "9":

μ = 0.774;



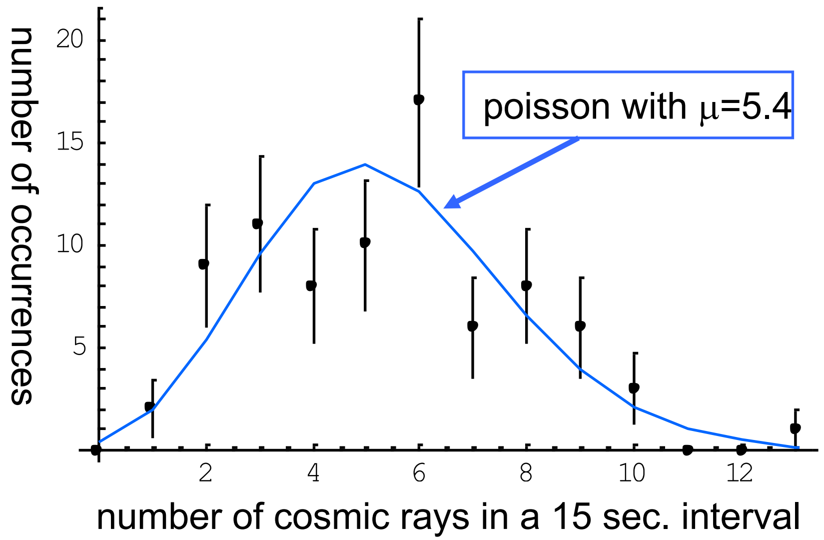
选读

Poisson Probability Distribution

About 10^{12} or 10^{13} total particles hitting the upper atmosphere per second.
A 1 cm^2 detector geometrically subtends $p \simeq 10^{-18}$ of the Earth's surface area

Counting the numbers of cosmic rays that pass through a detector in a 15 sec interval
Data is compared with a poisson using the measured average number of cosmic rays passing through the detector in eighty one 15 sec. intervals (μ=5.4)

counts	occurren-ces
0	0
1	2
2	9
3	11
4	8
5	10
6	17
7	6
8	8
9	6
10	3
11	0
12	0
13	1



Error bars are (usually) calculated using $\sqrt{n_i}$ (n_i=number in a bin)

Assume we have N total counts and the probability to fall in bin i is p_i.

For a given bin we have a binomial distribution (you're either in or out).

The expected average number in a given bin is: Np_i and the variance is $Np_i(1-p_i)=n_i(1-p_i)$

If we have a lot of bins then the probability of a event falling into a bin is small so $(1-p_i) \approx 1$

Distribution of time intervals

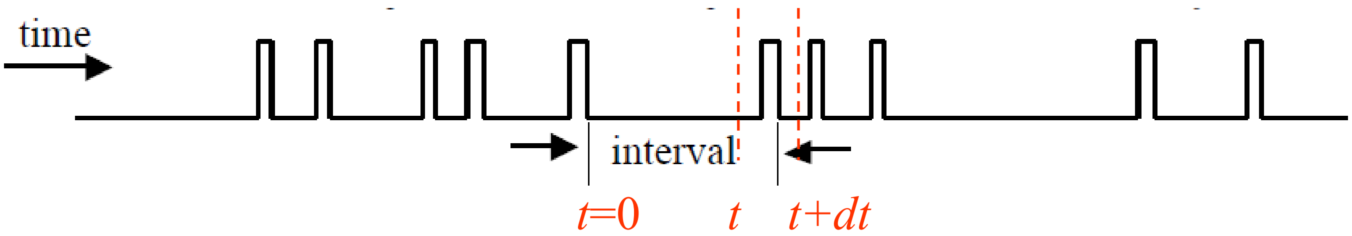
Poisson random process: random process charaterized by a <u>constant probability of occurence per unit time</u> regardless of past behavior

$dp = r\,dt$ $\quad r$: the average rate of occurence.

For the finite time interval T, the average number of events occurring will be rT

Intervals between Successive Events

Assume an event has occurred at time t=0 (*select a random point in time*).



probability of next event taking place in *dt* after delay of *t*	=	probability of *no* events during time from 0 to *t*	×	probability of an event during *dt*
$I_1(t)dt$	=	$P(0)$	×	$r\,dt$

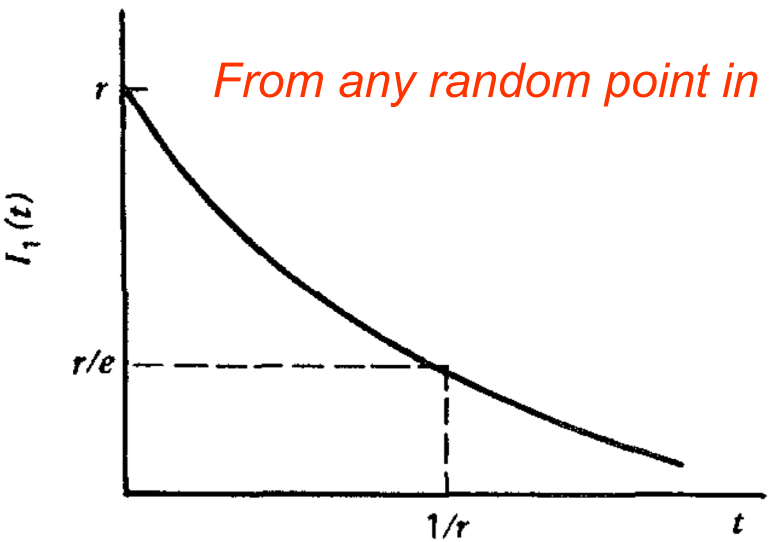
P(0): The probability that no event will recorded over an interval of length t for which the average number of recorded events should be *rt*

$$P(0) = \frac{(rt)^0 e^{-rt}}{0!}$$

$$P(0) = e^{-rt}$$

$$I_1(t)dt = re^{-rt}\,dt$$

From any random point in time



The most probable interval is zero.

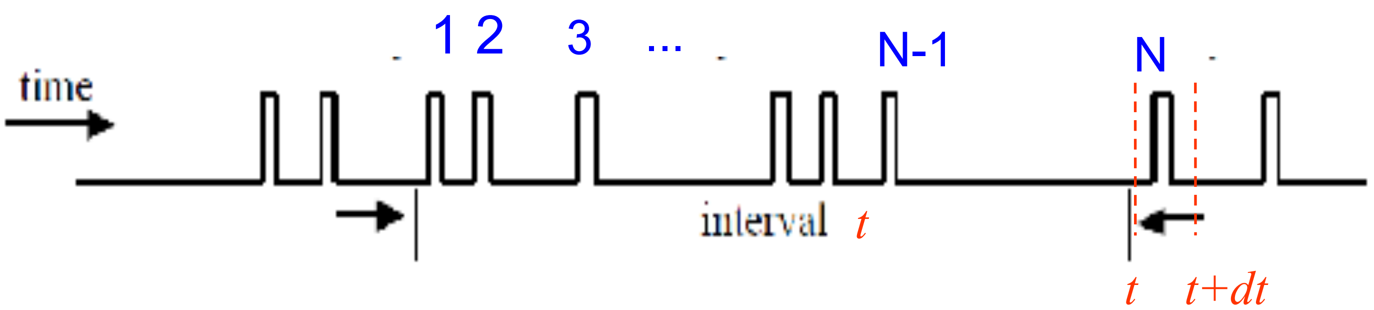
The average interval length

$$\bar{t} = \frac{\int_0^\infty tI_1(t)\,dt}{\int_0^\infty I_1(t)\,dt} = \frac{\int_0^\infty tre^{-rt}\,dt}{1} = \frac{1}{r}$$

Intervals Between Scaled Events

Events recorded by a digital scaler : a data buffer by producing an output pulse only when N input pulses have been accumulated.

"scale-down" high counting rates by a factor, N.



Two independent processes: A time interval of length t must be observed over which exactly N-1 events are presented to the scaler, and an additional event must occur in the increment dt following this time interval

$$I_N(t)\,dt = P(N-1)r\,dt$$

$$I_N(t)\,dt = \frac{(rt)^{N-1}e^{-rt}}{(N-1)!}\,r\,dt$$

$$\bar{t} = \frac{\int_0^\infty tI_N(t)\,dt}{\int_0^\infty I_N(t)\,dt} = \frac{N}{r}$$

$I_N(t)$ is the interval distribution for N-scaled intervals.

the most probable interval is evaluated by setting

$$\frac{dI_n(t)}{dt} = 0 \qquad t\Big|_{\text{most probable}} = \frac{N-1}{r}$$

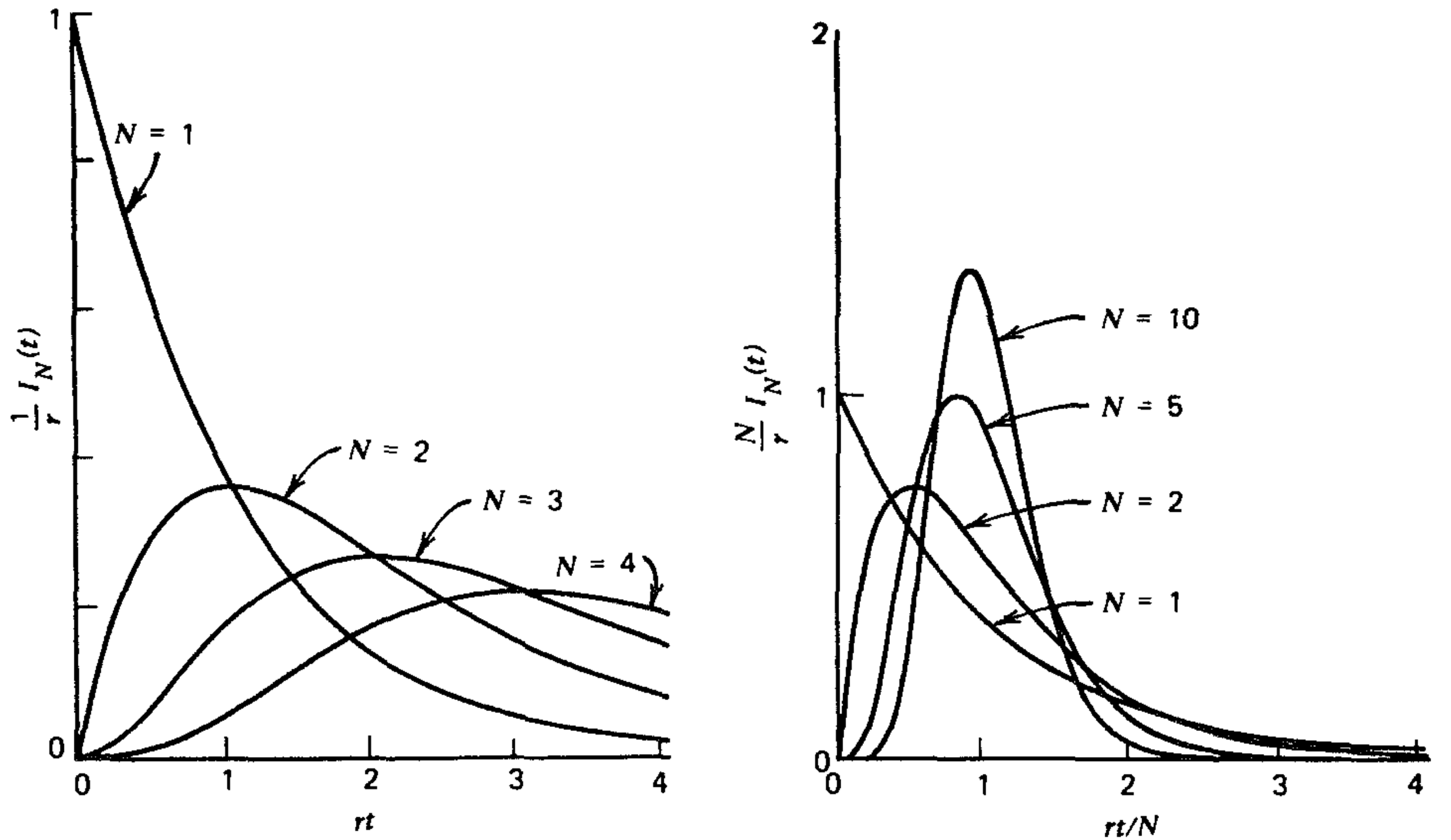
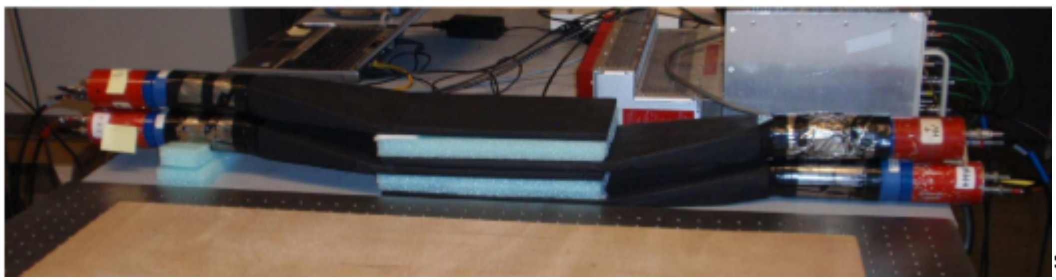
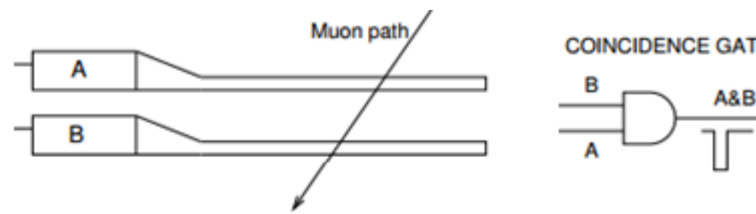


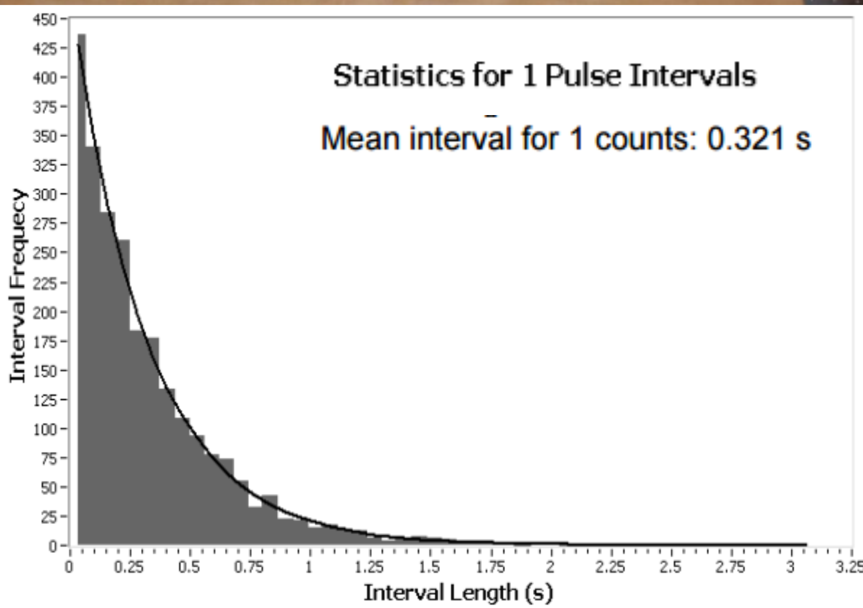
Figure 3.15 Graphical representation of the scaled interval distribution $I_N(t)$. (*a*) Four distributions for scaling factors of 1, 2, 3 and 4. (*b*) Interval distributions for $N = 1$ through $N = 10$ normalized to the same average interval N/r.

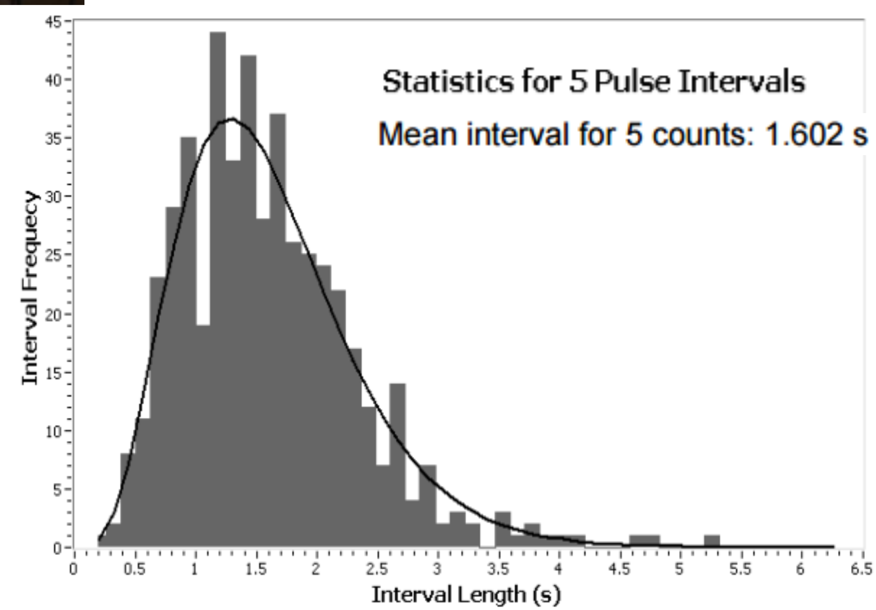
A Study of the Randomness of Cosmic Ray Arrival Times

It is conventional to assume that high energy cosmic rays are detected in the vicinity of the Earth at random times. Any deviations of massive particle primaries from random arrival distributions are expected only due to the lack of isotropy in the source distribution of particles such as might cause a time correlation on a diurnal basis.









http://www.publish.csiro.au/PH/pdf/PH800753

Null Experiments
Setting Confidence Limits When No Counts Are Observed

- Many experiments in physics test the validity of certain theoretical conservation laws by searching for the presence of specific reactions or decays forbidden by these laws. *For example, lifetime of proton*
- If no one or more events are observed within T, the theoretical law is disproven. However, if no events are observed, the converse cannot be said to be true. Instead a limit on the life-time of the reaction or decay is set.

For the process has some mean reaction rate *r* for N nuclei, the probability for observing no counts in a time period T is

$$P(0|r) = \exp(-rT)$$

$$P(0) = \frac{(rt)^0 e^{-rt}}{0!}$$

This can also be interpreted as the probability distribution for r when no counts are observed in a period *T.*

$$f(r) = T\exp(-rT)$$

$$\int_0^\infty f(r)dr = 1$$

T: nomalization factor

The probability that r is less r_0 is

$$P(r \leq r_0) = \int_0^{r_0} T \exp(-rT) dr$$
$$= 1 - \exp(-r_0 T)$$

This probability is known as *the confidence level(CL)* for the interval between 0 to r_0 $\quad CL = P(r \leq r_0) = 1 - \exp(-r_0 T)$

$$r_0 = -\frac{1}{T}\ln(1 - CL)$$ For N nuclei

$$r'_0 = -\frac{1}{NT}\ln(1 - CL)$$ For each nucleus

$r \leq r_0$ is with CL confidence level. (upper limit)

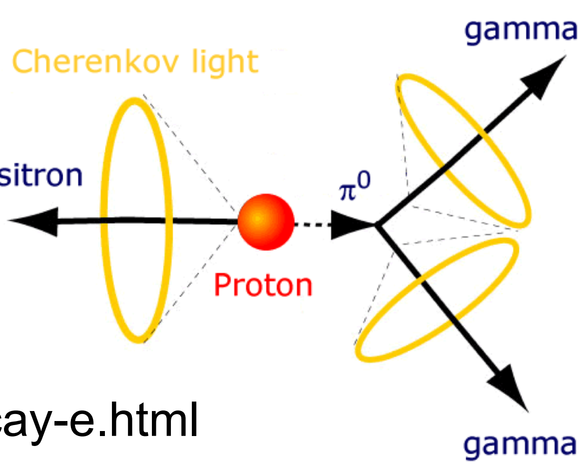
$$\tau = \frac{1}{r} \geq -\frac{NT}{\ln(1 - CL)}$$ Mean lifetime

To make a strong statement we can choose a high confidence level (CL), for example, 90%.

Proton decay measurement at the **Super-Kamiokande**

Super-Kamiokande uses 50,000 tons of pure water and it contains 7×10^{33} protons. Super-Kamiokande has started measurement since 1996 and is running more than 10 year, however, we have not observed any evidence of proton decay yet. From this result, proton lifetime is estimated to be more than 10^{34} years(at 90%CL) (age of the universe ~10^{10} years).

$$\tau \geq -\frac{NT}{\ln(1-CL)} = \frac{7 \times 10^{33} \times 10}{2.303} \; years$$



http://www-sk.icrr.u-tokyo.ac.jp/sk/sk/pdecay-e.html

选读

Example 4.4 A 50 g sample of ^{82}Se is observed for 100 days for neutrinoless double beta decay, a reaction normally forbidden by lepton conservation. However, current theories suggest that this might occur. The apparatus has a detection efficiency of 20%. No events with the correct signature for this decay are observed. Set an upper limit on the lifetime for this decay mode.

Choosing a confidence limit of 90%, (4.59) yields

$$\lambda \le \lambda_0 = -\frac{1}{100 \times 0.2} \ln(1 - 0.9) = 0.115 \text{ day}^{-1},$$

where we have corrected for the 20% efficiency of the detector. This limit must now be translated into a lifetime per nucleus. For 50 g, the total number of nuclei is

$$N = \frac{N_a}{82} \times 50 = 3.67 \times 10^{23},$$

which implies a limit on the decay rate per nucleus of

$$\lambda \le \frac{0.115}{3.67 \times 10^{23}} = 3.13 \times 10^{-25} \text{ day}^{-1}.$$

The lifetime is just the inverse of λ which yields

$$\tau \ge 8.75 \times 10^{21} \text{ years} \qquad 90\% \text{ CL},$$

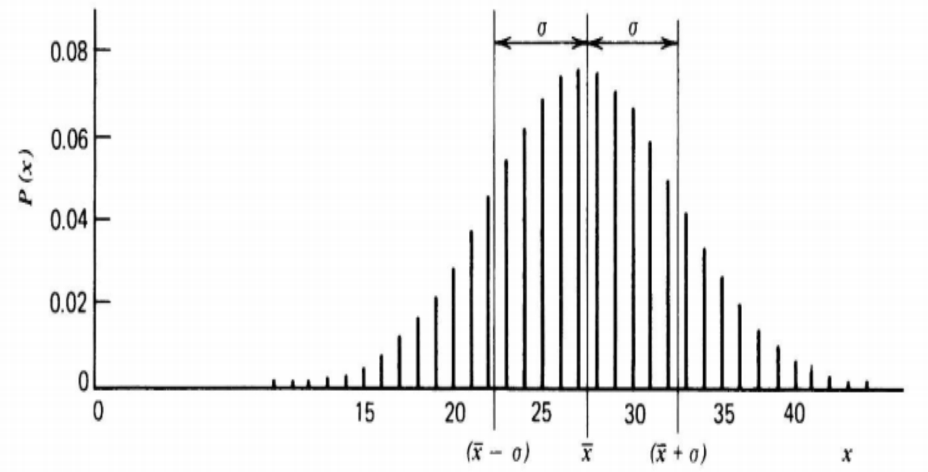
where we have converted the units to years. Thus, neutrinoless double beta decay may exist but it is certainly a rare process!

The Gaussian or Normal Distribution

The Gaussian distribution plays a central role in all of statistics and is the most ubiquitous distribution in all the science.

Measurements errors, and in particular, instrumental errors are generally discribed by this probability distribution.

$$P(r) = \frac{1}{\sqrt{2\pi\mu}} \exp(-\frac{(r-\mu)^2}{2\mu})$$



μ = average of the distribution

$\sigma^2 = \mu$

In general, probability density function $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

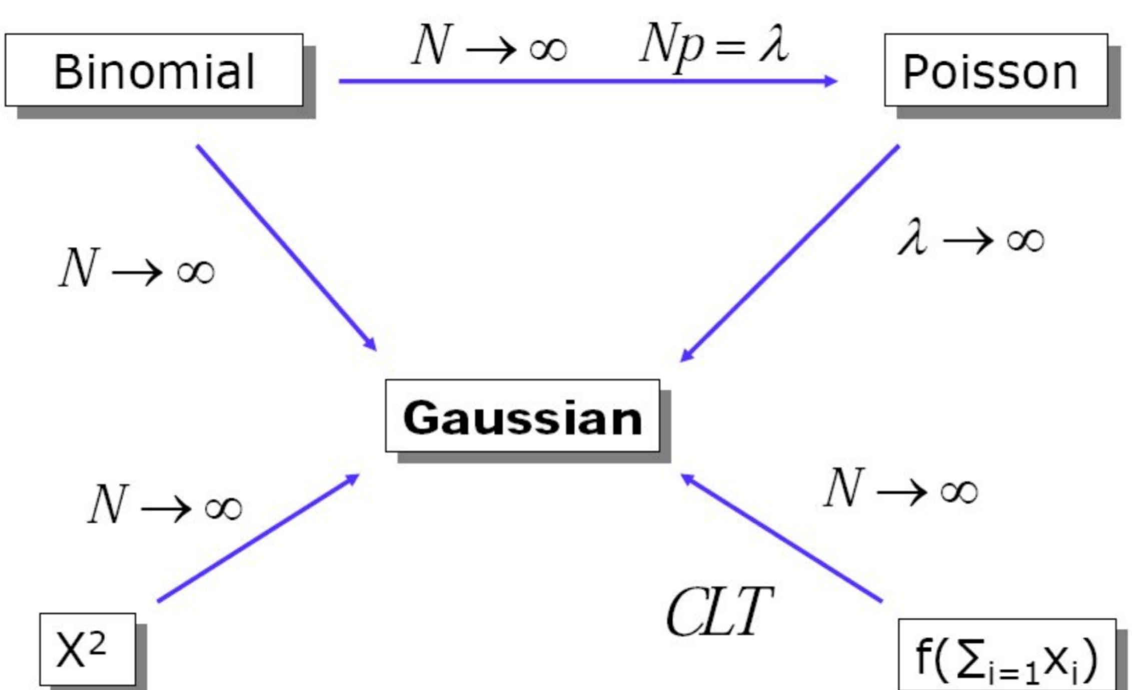
When μ is large, the Poisson distribution $P_\mu(\nu)$ is well approximated by the Gauss function with the same mean and standard deviation

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

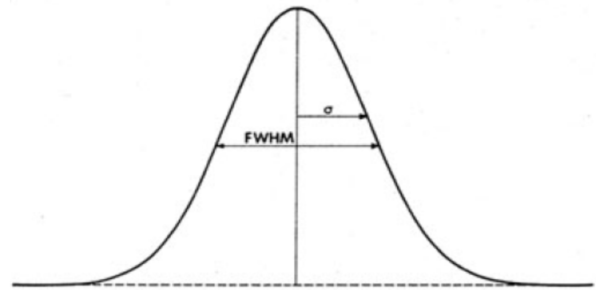
which by make a suitable coordinate transformation, $x \to \sigma x + \mu$, gives the **Normal distribution**

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

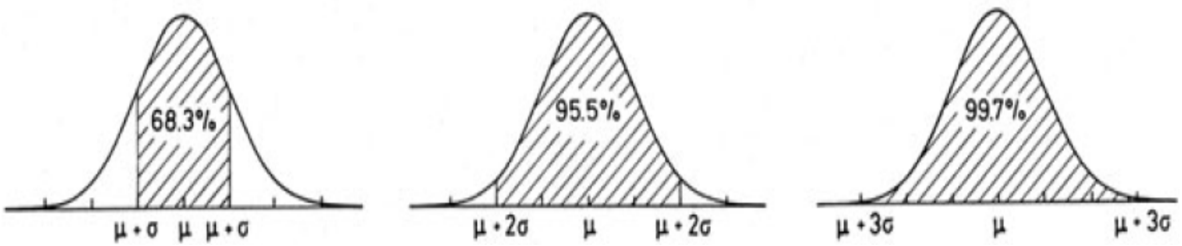
Mean = zero
Rms = 1



The full width at half maximum (FWHM)



$$FWHM = 2\sigma\sqrt{2\ln 2} = 2.35\sigma$$



The area contained between the limits $\mu\pm1\sigma$, $\mu\pm2\sigma$ and $\mu\pm3\sigma$ in a Gaussian distribution.

The presentation of a result as $x\pm\sigma$ signifies, in fact, that the true value has 68% probability of lying between the limits $x-\sigma$ and $x+\sigma$

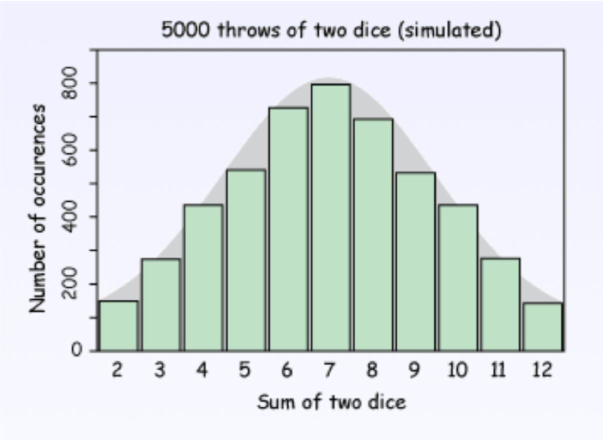
Gaussian and CLT

Central Limit Theorem (CLT):

- Consider the sum of X of N independent variables x_i, with i = 1,2,3... , each taken from a different distribution with mean μ_i and variance σ_i^2

- Then the distribution for $X = \sum x_i$ has the following properties:
  – its expectation value is $\mu = \sum \mu_i$
  – its variance is $\sigma^2 = \sum \sigma_i^2$
  – it becomes Gaussian distributed for $n \to \infty$

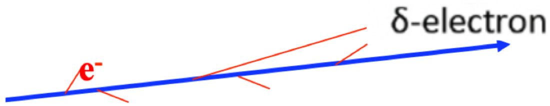
- For CLT to be valid:
  - μ and σ of *pdf* must be finite.
  - No one term in sum should dominate the sum.



This is highly relevant for experimental resolutions (see later lecture), because many different sources for errors in measurements are mostly independent

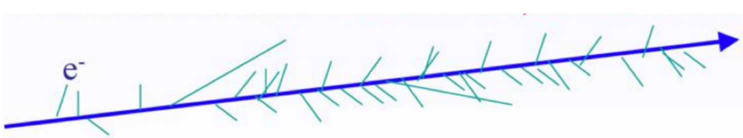
Energy straggling: The energy loss distribution

- The distribution for thin absorbers:



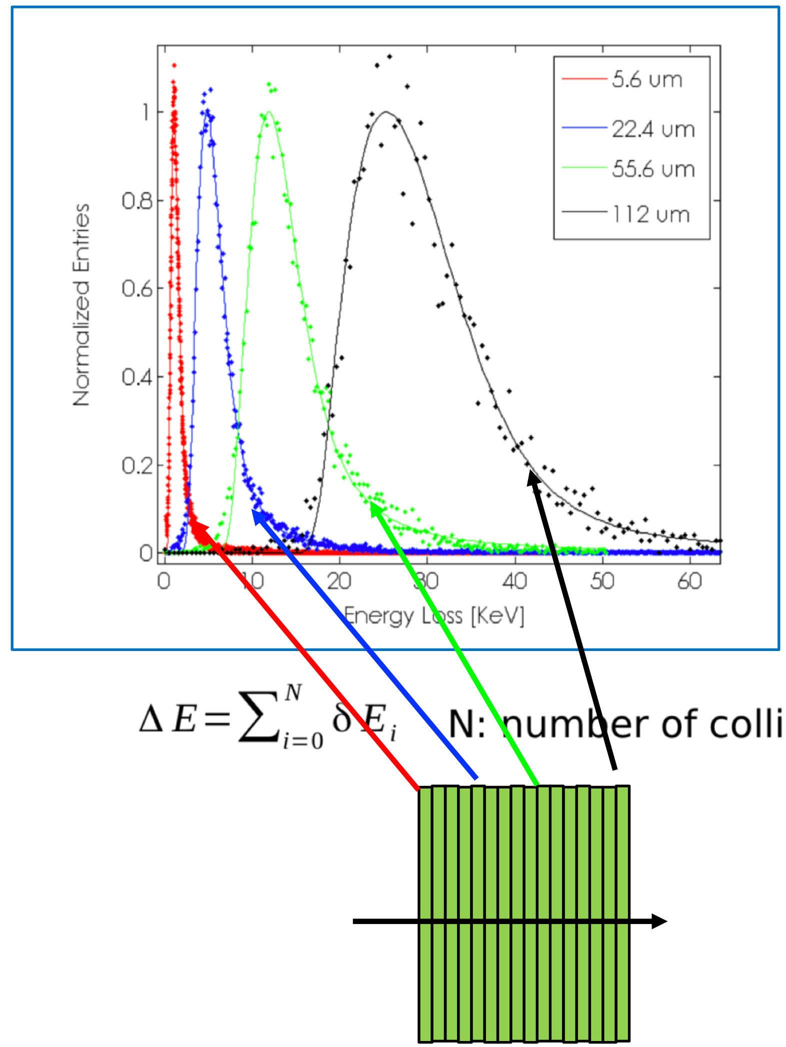
  - ➢ few collisions, some with high energy transfer(δ-electrons).
  - ➢ Energy loss distributions shows large fluctuation towards high losses: Landau distributions

- For relatively thick absorbers:



  - ➢ Many collisions, the total energy loss follows directly from the *Central Limit Theorem and* approches the Gaussian form

Energy loss distributions for 12GeV protons passing through several silicon thicknesses.



$$\Delta E = \sum_{i=0}^{N} \delta E_i$$ N: number of collisions

E=NδE, δE →0,N →∞

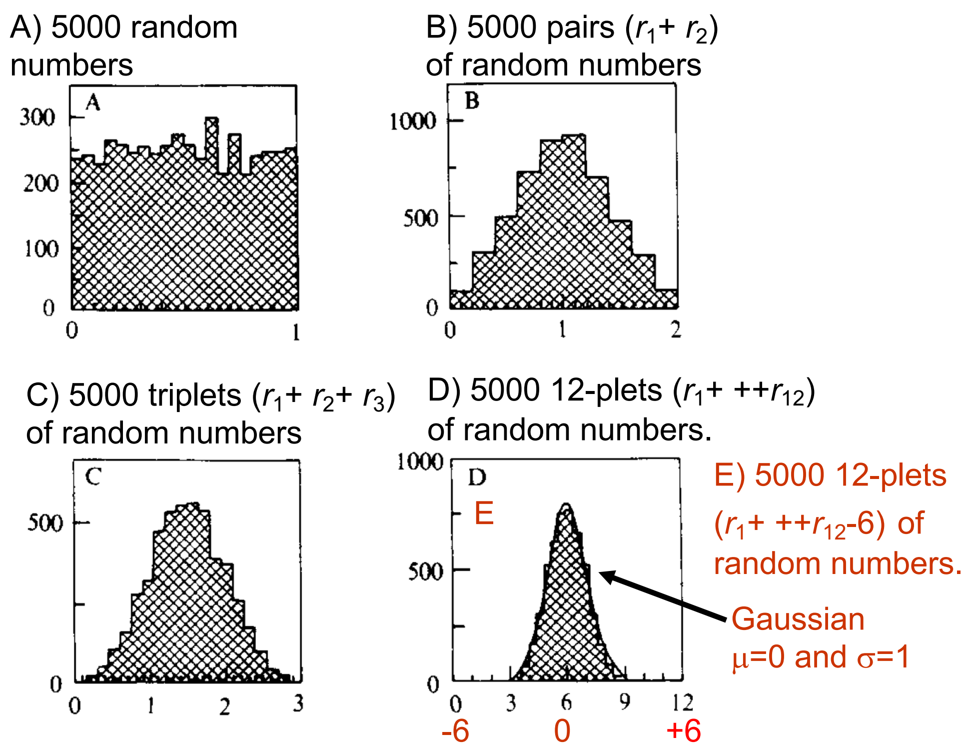
Best illustration of the CLT.

a) Take 12 numbers (r_i) from your computer's random number generator
b) add them together
c) Subtract 6
d) get a number that is from a gaussian *pdf* !

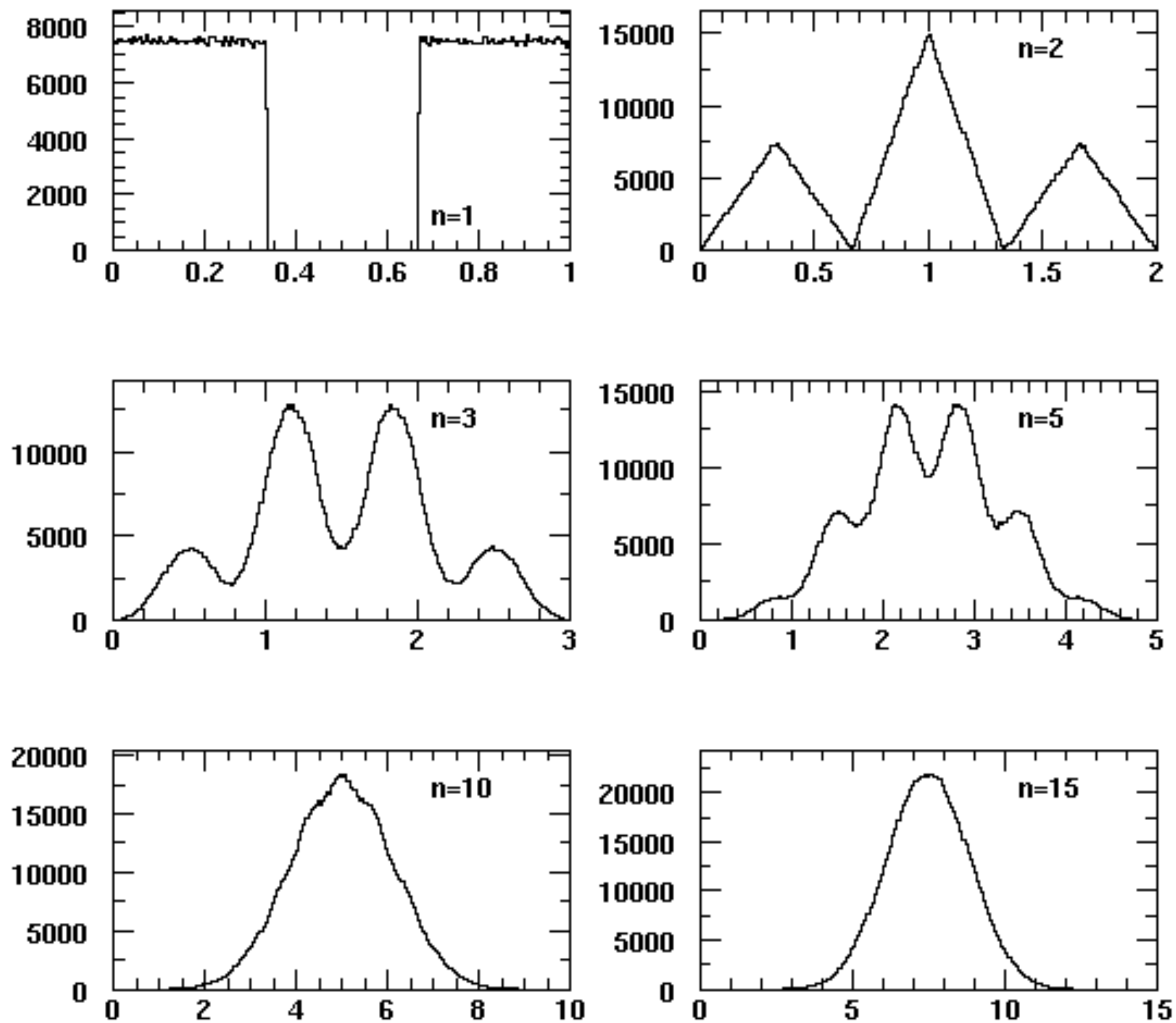
Computer's random number generator gives numbers distributed uniformly in the interval [0,1].

A uniform *pdf* in the interval [0,1] has $\mu=1/2$ and $\sigma^2=1/12$

Thus the sum of 12 uniform random numbers minus 6 is distributed as if it came from a gaussian *pdf* with $\mu=0$ and $\sigma=1$.



In this case, 12 is close to ∞.



Repeated measurements will give a normal distribution about the mean

Measurement Errors (Uncentainties)

Statistical Errors

Differences in results are randomly varying, giving *statistical uncertainties ,*
There is the law of large numbers applies and helps to increase precision !

Statistical error is usually assumed to be from a Gaussian distribution.
With the assumption of Gaussian statistics we can say (calculate) something about how well our experiment agrees with other experiments and/or theories.

Expect ~ 68% chance that the true value is between x - σ and x + σ .

The error in the mean s_m:
If we repeat a measurement n times and each measurement has uncertainty s，then

$$\sigma_m = \frac{\sigma}{\sqrt{n}} \to 0$$

We will see it later

Systematic Errors

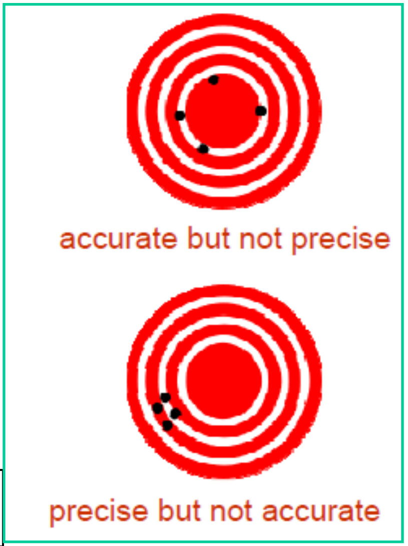
A *systematic error* denotes the uncertainty in estimating effects caused by systematic mistakes and caused by neglecting systematic mistakes.

Systematic mistakes are eg. wrong method, faulty instruments, wrong formulae ..

Because of systematic errors, an experimental result can be precise, but not accurate!

Comments on systematic uncertainties:

- sys. errors do NOT decrease with $1/\sqrt{n}$
- statistical and systematic errors are in general independent.
- need to quote errors separately in the results.



x = 10.2 ± 0.2 (statistical) ± 0.3 (systematic) ± 0.3 (theory) [units]

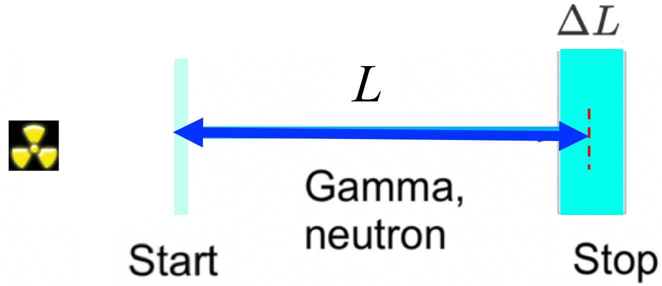
Often errors are NOT symmetric.
Need to quote both errors:

$$value = x_0 \, {}^{+\sigma_{up}}_{-\sigma_{low}}$$

Propagation of errors

Experiment observable x : $\bar{x} \pm \sigma_x$

Derived quantity $f(x)$: $f(\bar{x}) \pm \sigma_f(?)$



$$L, TOF \rightarrow E_n$$

$$\Delta L, \Delta t \rightarrow \Delta TOF \rightarrow \Delta E_n$$

$$E_n = \frac{1}{2} m_n (\frac{L}{TOF})^2$$

$$E_n \pm \Delta En$$

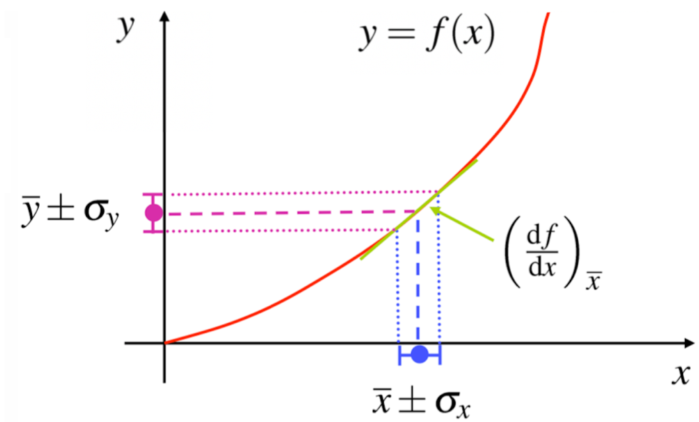
When σ_x is enough to use linear approximation to f[x] near $\bar{x}$,

$$f - \bar{f} \approx \frac{\partial f}{\partial x}(x - \bar{x})$$

$$\sigma_f^2 = \frac{1}{N-1}\sum_{i=1}^{N}(f_i - \bar{f})^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[\frac{\partial f}{\partial x}(x_i - \bar{x})\right]^2$$

$$\sigma_f^2 = (\frac{\partial f}{\partial x})^2 \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

$$\boxed{\sigma_f^2 = (\frac{\partial f}{\partial x})^2 \sigma_x^2}$$



Experiment observables:

$$x : \bar{x} \pm \sigma_x$$
$$y : \bar{y} \pm \sigma_y$$

→

Derived quantity

$$f(x,y) : f(\bar{x},\bar{y}) \pm \sigma_f (?)$$

$$f - \bar{f} \approx \frac{\partial f}{\partial x}(x - \bar{x}) + \frac{\partial f}{\partial y}(y - \bar{y})$$

$$\sigma_f^2 = \frac{1}{N-1}\sum_{i=1}^{N}(f_i - \bar{f})^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[\frac{\partial f}{\partial x}(x_i - \bar{x}) + \frac{\partial f}{\partial y}(y_i - \bar{y})\right]^2$$

$$\sigma_f^2 = \frac{1}{N-1}\left[\left(\frac{\partial f}{\partial x}\right)^2 \sum_{i=1}^{N}(x_i - \bar{x})^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sum_{i=1}^{N}(y_i - \bar{y})^2 + 2\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})\right]$$

$$\sigma_{x,y} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$ is their *covariance*. 协方差

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

Covariance and correlation

协方差 关联

Define covariance of x and y ,cov[x,y] as

$$cov[x,y] = E[xy] - \mu_x \mu_y = E[(x-\mu_x)(y-\mu_y)]$$

For $f(x,y,z)$, there are three covariances: cov(x,y), cov(x,z), cov(y,z).

Correlation coefficient (dimensionless) defined as

$$\rho_{xy} = \frac{cov[x,y]}{\sigma_x \sigma_y}$$

ρ ranges between +1 and -1

If x, y, independent, i.e., $f(x,y) = f_x(x) f_y(y)$, then

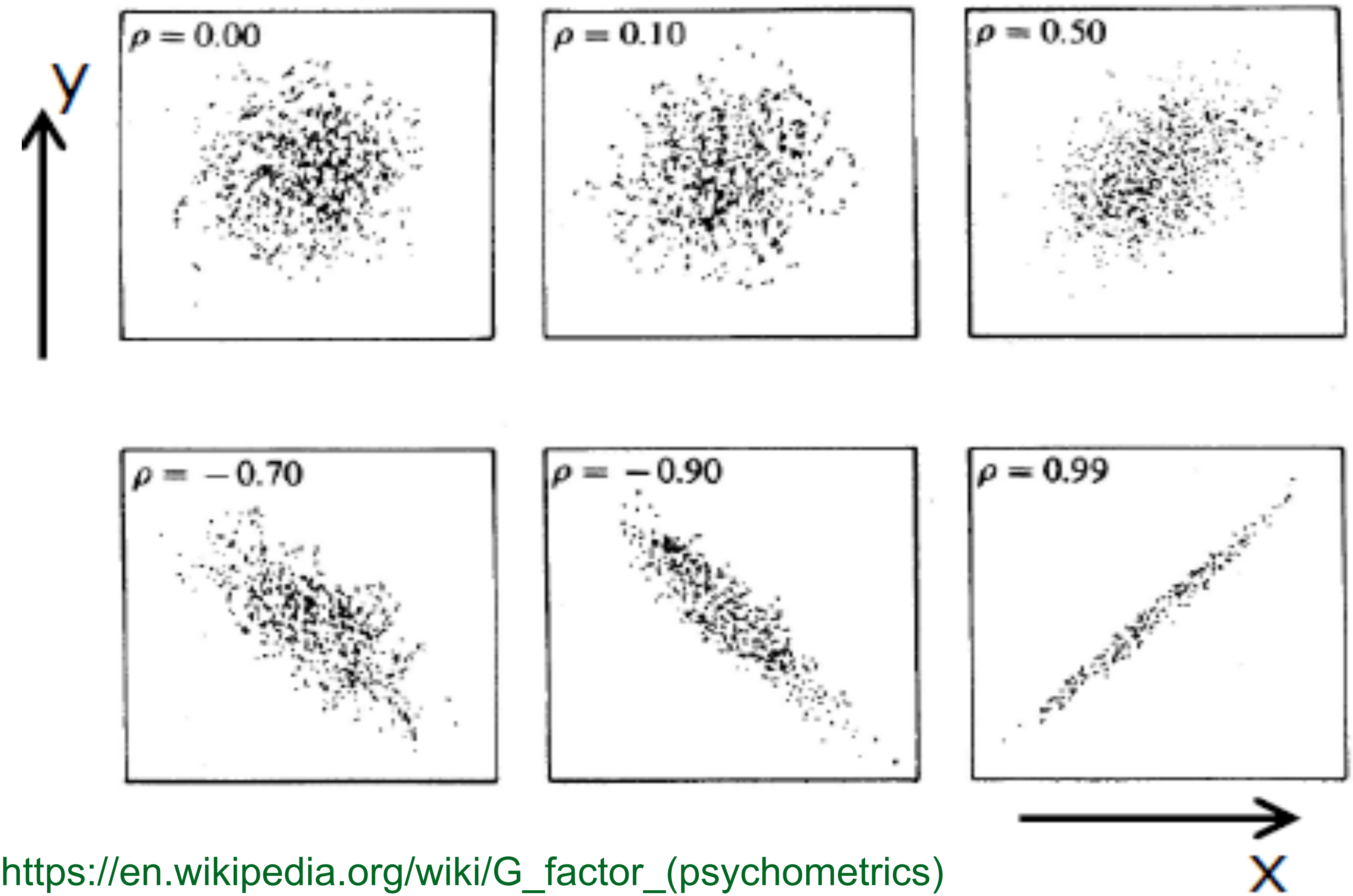
$$E[xy] = \int \int xy\, f(x,y)\, dx dy = \mu_x \mu_y$$

$\rightarrow \rho_{xy} = 0$ x and y, ‘uncorrelated’

If x, y, correlated linealy, then $|\rho_{xy}|=1$

Correlation

ρ=+1 if and only if Y =aX+b with a>0,
ρ=−1 if and only if Y =aX+b with a<0.



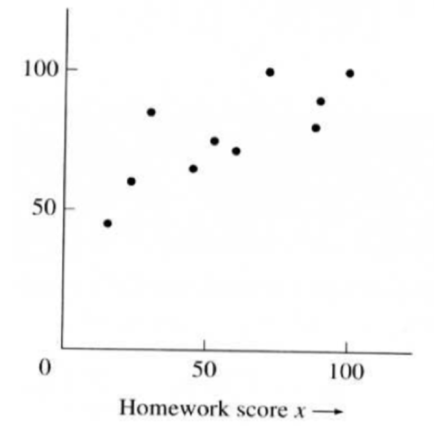
https://en.wikipedia.org/wiki/G_factor_(psychometrics)

http://www.psych.utoronto.ca/users/reingold/courses/intelligence/cache/1198gottfred.html

As an example, consider the exam and homework scores shown in Figure 9.1. These scores are given in Table 9.3. A simple calculation (Problem 9.12) shows that

Table 9.3. Students' scores.

Student i	1	2	3	4	5	6	7	8	9	10
Homework x_i	90	60	45	100	15	23	52	30	71	88
Exam y_i	90	71	65	100	45	60	75	85	100	80



the correlation coefficient for these 10 pairs of scores is $\rho = 0.8$. The professor concludes that this value is "reasonably close" to 1 and so can announce to next year's class that, because homework and exam scores show good correlation, it is important to do the homework.

If our professor had found a correlation coefficient ρ close to zero, he would have been in the embarrassing position of having shown that homework scores have no bearing on exam scores. If ρ had turned out to be close to -1, then he would have made the even more embarrassing discovery that homework and exam scores show a *negative* correlation; that is, that students who do a good job on homework tend to do poorly on the exam.

From John R. Taylor, An introduction to error analysis 2nd edition

Propagation of errors

Error estimates for a function of many correlated variables $f(x_1, x_2, \ldots x_n)$, need to take correlation into account:

$$\sigma_f^2 = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)^2 \cdot \sigma_{x_i}^2 + \sum_{i=1}^{n} \sum_{j \neq i}^{n} \left(\frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \right) \cdot \mathrm{cov}(x_i, x_j)$$

Normal errors (for uncorrelated variables).

Additional terms accounting for correlations

If x_1, x_2, . . . , x_n are independent quantities having errors σ_{x1}, σ_{x2}, .σ_{xn}

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x_1} \right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2} \right)^2 \sigma_{x_2}^2 + \ldots + \left(\frac{\partial f}{\partial x_n} \right)^2 \sigma_{x_n}^2$$

Propagation of Uncertainty (neglecting correlations)

$$f = x + y \rightarrow \sigma_f^2 = \sigma_x^2 + \sigma_y^2$$

$$f = x - y \rightarrow \sigma_f^2 = \sigma_x^2 + \sigma_y^2$$

Errors add in "quadrature"

$$f = x \times y \rightarrow (\frac{\sigma_f}{f})^2 = (\frac{\sigma_x}{x})^2 + (\frac{\sigma_y}{y})^2$$

$$f = x/y \rightarrow (\frac{\sigma_f}{f})^2 = (\frac{\sigma_x}{x})^2 + (\frac{\sigma_y}{y})^2$$

Relative Errors add in "quadrature"

$$f = x^n \rightarrow \frac{\sigma_f}{f} = n\frac{\sigma_x}{x}$$

Only one measurement, estimate uncertainty?

There is only one measurement, strictly speaking you are out of luck.
However, one can posit that it must be the mean, and σ^2 = mean (Poisson distribution)
One further assumes that the distribution is symmetric: $\bar{x} \pm \sigma$

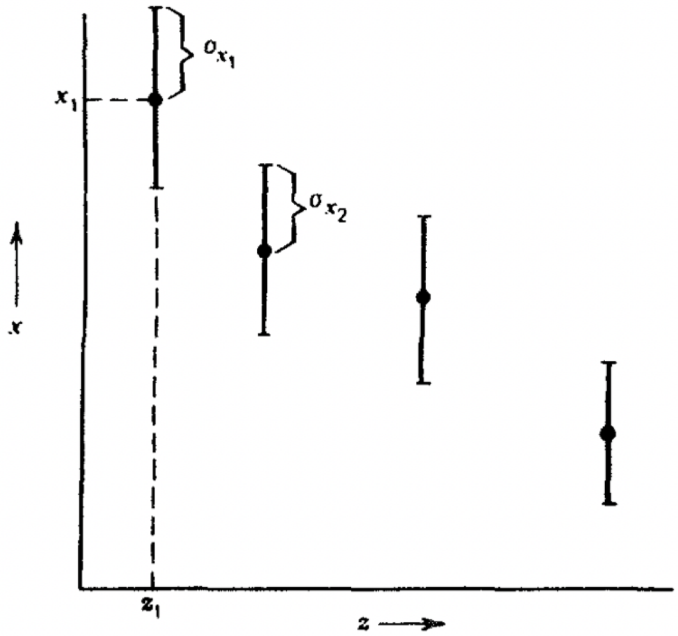


Figure 3.13 A graphical display of error bars associated with experimental data.

Uncertainty of Mean of N measurements

N *independent measurements* of the same physical quantity:

$$x_1, x_2, x_3, \ldots, x_i, \ldots, x_N \qquad \sigma_x = \sigma_{x_1} = \sigma_{x_2} \ldots = \sigma_{x_n}$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\sigma_{\bar{x}}^2 = \sum_{i=1}^{N} \left(\frac{\partial \bar{x}}{\partial x_i}\right)^2 \sigma_{x_i}^2 = \sum_{i=1}^{N} \left(\frac{\sigma_{x_i}}{N}\right)^2 = \frac{\sigma_x^2}{N}$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}}$$

The uncertainty in the mean is smaller than the uncertainty in a single observation by a factor of $1/\sqrt{N}$

Example 4.3 Consider the following series of measurements of the counts per minute from a detector viewing a ^{22}Na source,

2201	2145	2222	2160	2300

What is the decay rate and its uncertainty?

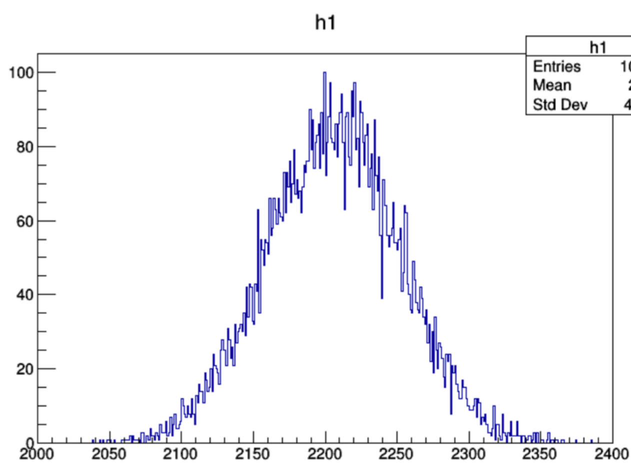
Since radioactive decay is described by a Poisson distribution, we use the estimators for this distribution to find

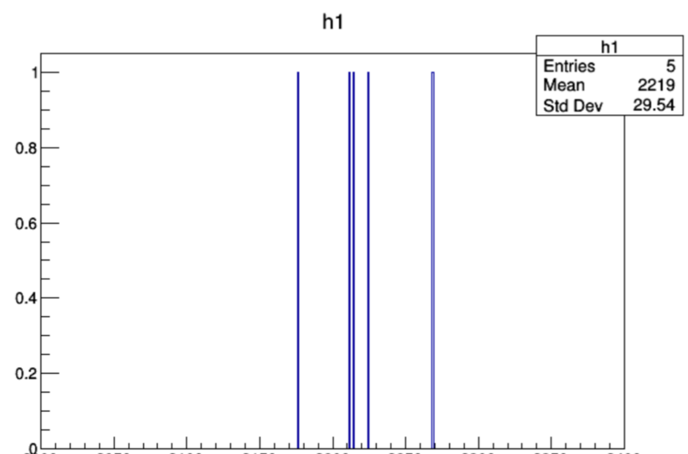
$$\hat{\mu} = \bar{x} = 2205.6 \quad \text{and}$$

$$\sigma(\hat{\mu}) = \sqrt{\frac{\bar{x}}{n}} = \sqrt{\frac{2205.6}{5}} = 21 \, .$$

The count rate is thus

$$\text{Count Rate} = (2206 \pm 21) \text{ counts/min.}$$

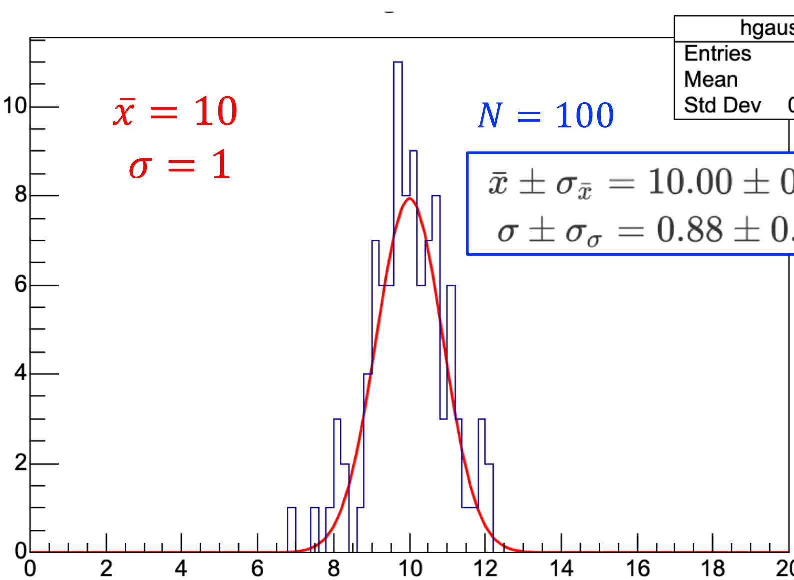


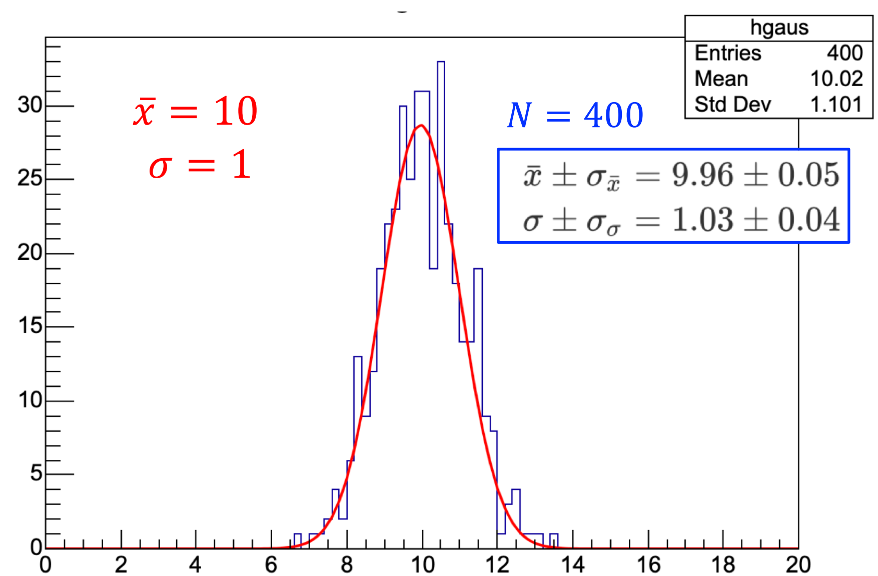


- Do not confuse $\sigma_{\bar{x}}$ with σ !
  - σ is related to the width of the pdf that measurements come from. -resolution
  - σ does not get smaller as we combine measurements.

Example: x follows gaussian distribution:

$$G(x,\bar{x},\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \qquad \sigma_{\bar{x}} = \sigma/\sqrt{N}$$





Combination of independent Measurements with unequal errors

N independent measurements with different uncertainty.

Idea: give more weight to those measurements with small values of σ_{xi} and less weight to measurements for which this estimated error is large.

- Let each individual measurement x_i be given a weighting factor a_i and the best value $\bar{x}$ computed from the linear combination

$$\bar{x} = \frac{1}{\alpha}\sum_{i=1}^{N} a_i x_i \qquad \alpha = \sum_{i=1}^{N} a_i$$

$$\sigma_{\bar{x}}^2 = \sum_{i=1}^{N} \left(\frac{\partial \bar{x}}{\partial x_i}\right)^2 \sigma_{x_i}^2 = \frac{1}{\alpha^2}\sum_{i=1}^{N} a_i^2 \sigma_{x_i}^2$$

The weighting factor a_i should be chosen in order to minimize the expected error in $\bar{x}$.

$$\frac{\partial \sigma_{\bar{x}}^2}{\partial a_j} = 0 \rightarrow a_j = \frac{1}{\sigma_{x_j}^2} \left(\sum_{j=1}^{N} \frac{1}{\sigma_{x_j}^2} \right)^{-1}$$

$$\bar{x} = \frac{1}{\sum 1/\sigma_i^2} \sum x_i/\sigma_i^2$$

The weighted mean

$$\sigma_{\bar{x}}^2 = \sum_i \frac{1}{1/\sigma_i^2}$$

The error on the weighted mean

Example 4.2 It is necessary to use the lifetime of the muon in a calculation. However, in searching through the literature, 7 values are found from different experiments:

2.198 ± 0.001 μs	2.203 ± 0.004 μs	2.202 ± 0.003 μs
2.197 ± 0.005 μs	2.198 ± 0.002 μs	2.1966 ± 0.0020 μs
2.1948 ± 0.0010 μs		

What is the best value to use?

One way to solve this problem is to take the measurement with the smallest error; however, there is no reason for ignoring the results of the other measurements. Indeed, even though the other experiments are less precise, they still contain valid information on the lifetime of the muon. To take into account all available information we must take the weighted mean. This then yields then mean value

$$\tau = 2.19696$$

with an error

$$\sigma(\tau) = 0.00061 .$$

Note that this value is smaller than the error on any of the individual measurements. The best value for the lifetime is thus

$$\tau = 2.1970 \pm 0.0006 \ \mu s \ .$$

Optimization of counting experiments

$n_s=N_s/t_s$: counting rate due to source and background

$n_b=N_b/t_b$: counting rate due to background

$n_0=n_s-n_b$: counting rate due to source alone

$t=t_s+t_b$

- For a given $t=t_s+t_b$, uncertanty of n_0 can be minimized by optimally choosing the fraction of t allocated to t_s(or t_b)

$$\sigma_n = \sqrt{\sigma_N^2 / t^2} = \sqrt{N / t^2} = \sqrt{n / t}$$

$$\sigma_{n_0} = \sqrt{n_s / t_s + n_b / t_b} = \sqrt{n_s / t_s + n_b /(t - t_s)}$$

$$\frac{\partial \sigma_{n_0}}{\partial t_s} = 0 \Rightarrow \quad t_s / t_b \Big|_{opt} = \sqrt{n_s / n_b}$$

$$\Rightarrow v_{n_0 \min} = \sigma_{n_0} / n_0 \Big|_{\min} = \frac{1}{(\sqrt{n_s} - \sqrt{n_b})} \frac{1}{\sqrt{t}} \qquad = \frac{\sqrt{n_b}(\sqrt{n_s / n_b} + 1)}{n_0} \frac{1}{\sqrt{t}}$$

For low level measurement $n_b \sim n_s$, i.e. $n_s/n_b \sim 1$

n_0 is proportion to detection efficiency ε

$$v_{n_0 \min} \propto 1/(\varepsilon / \sqrt{n_b})$$

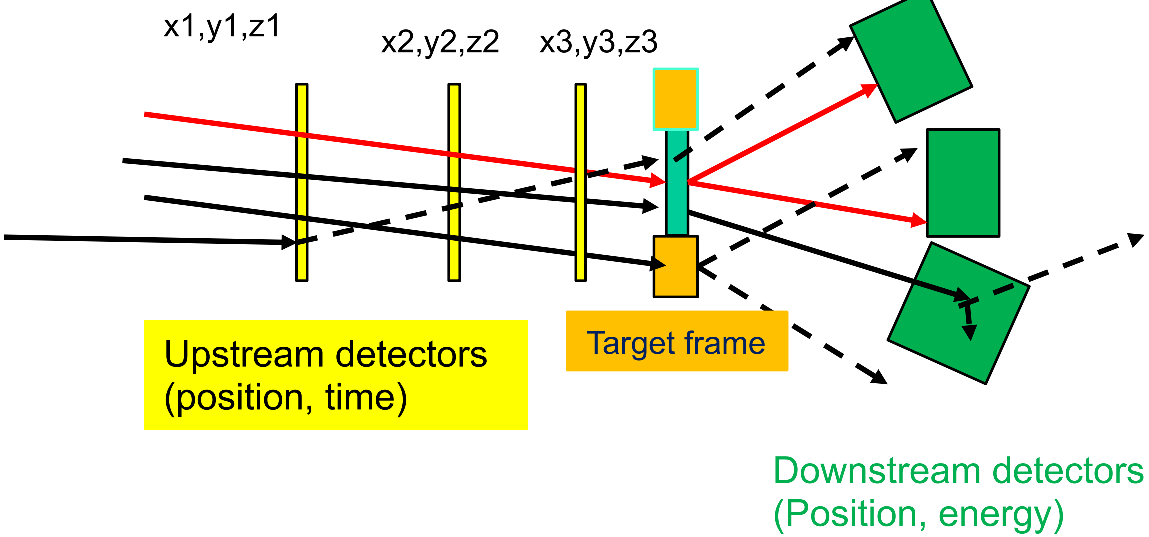
It's vital to use a detector with high detection efficiency and low background for low level measurement

For a given ν_{n0min}

$$t_{\min} = \frac{1}{(\sqrt{n_s} - \sqrt{n_b})} \frac{1}{v_{n_0 \min}}$$

Background run:

Experiment with target out to estimate the reactions in detectors and target frame



Y. L. YE *et al.*

PHYSICAL REVIEW C **71**, 014604 (2005)

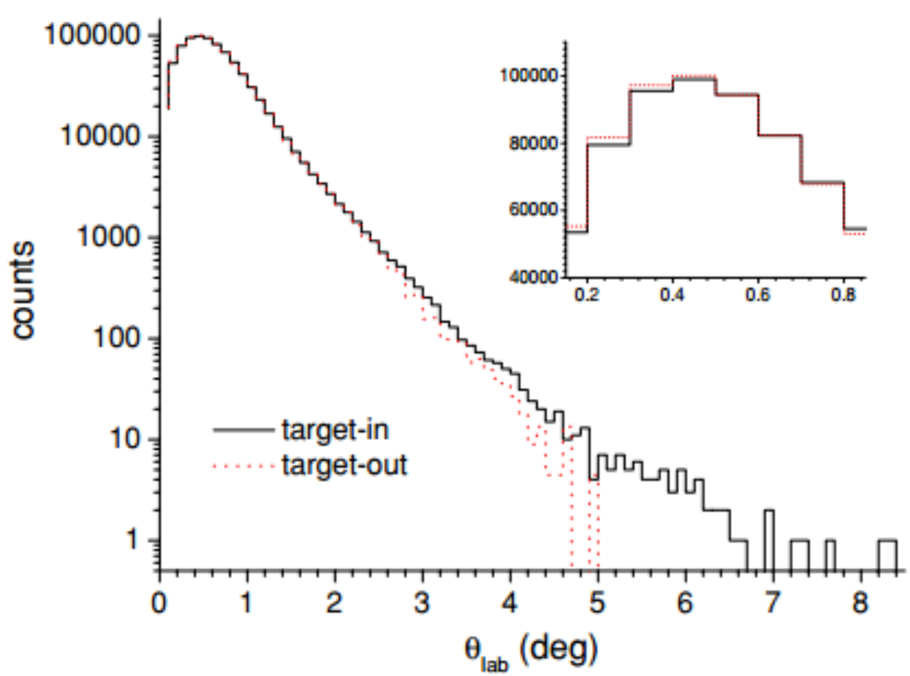
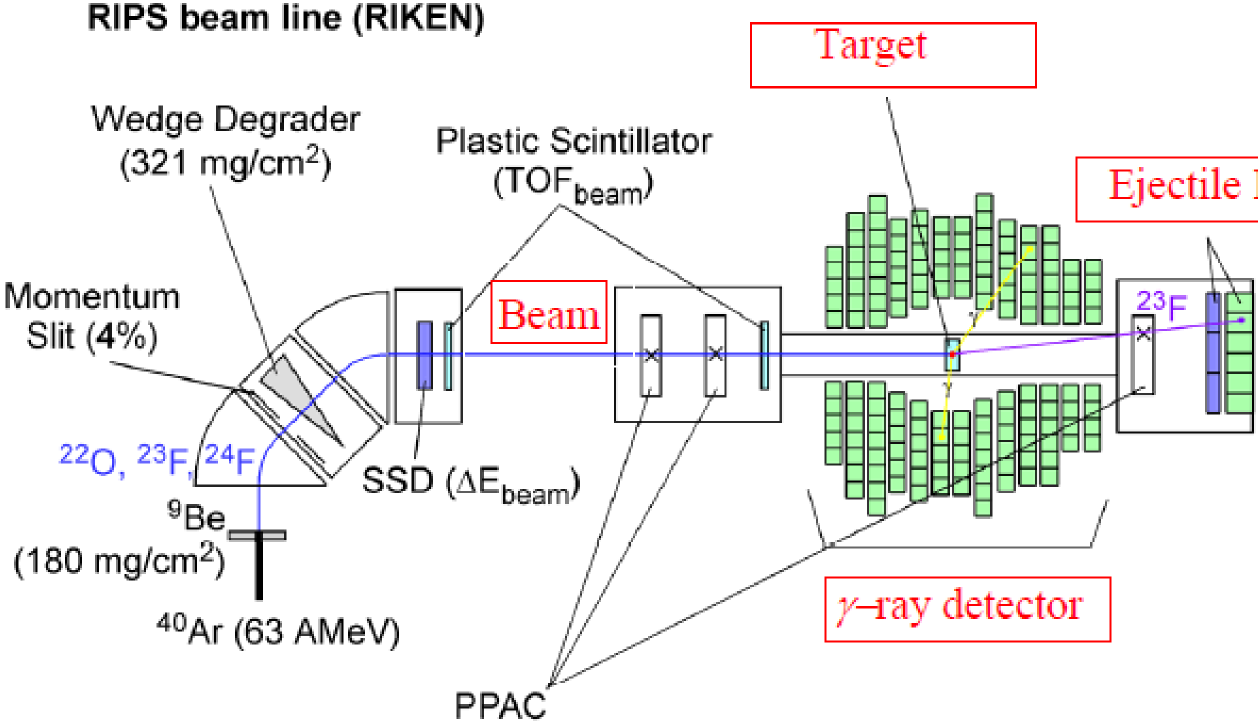
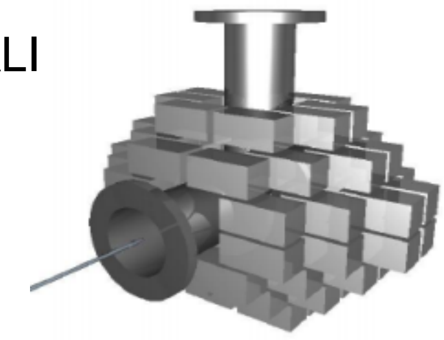


FIG. 4. (Color online) Angular distribution of the ^{6}He particles for target-in (solid line) and target-out (dotted line) runs, detected by the 0° telescope. The inset is a linear display of the counts at very small angles.

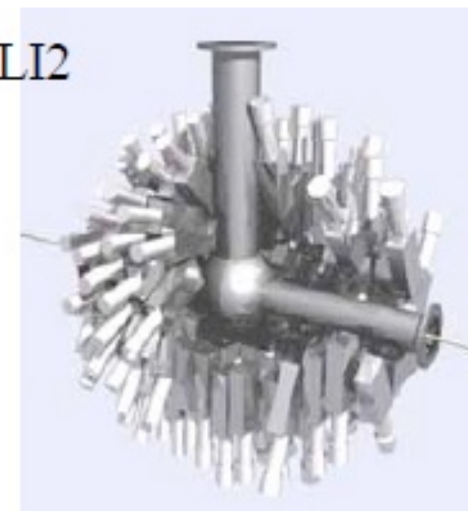
Example: In-beam gamma spectroscopy



DALI



DALI2

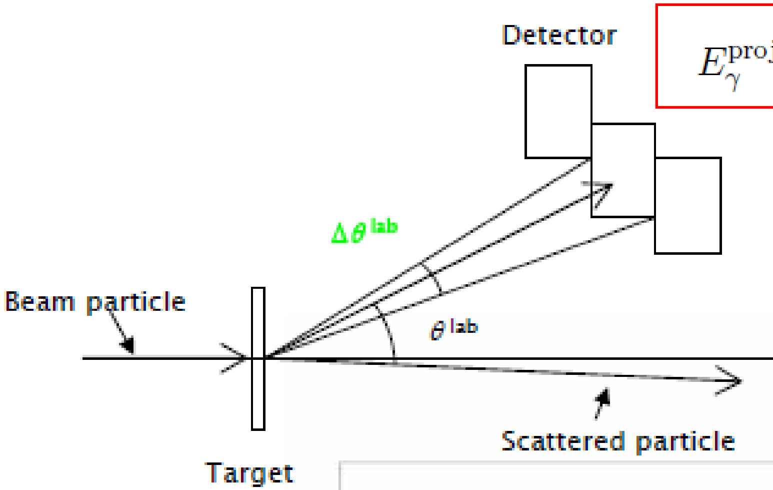


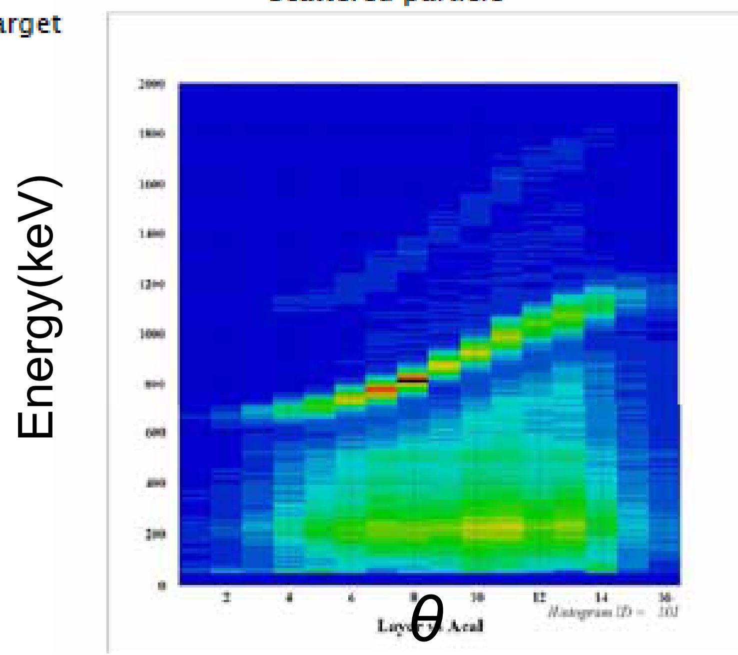
# of NaI(Tl) detectors	160
Angular resolution	8 deg.
Efficiency @ 1MeV	21%
E resol. (β=0.3) @1MeV	8%
Coverage (θ)	30~160 deg.

Doppler correction of γ-rays from fast RI Beam (β~0.3)

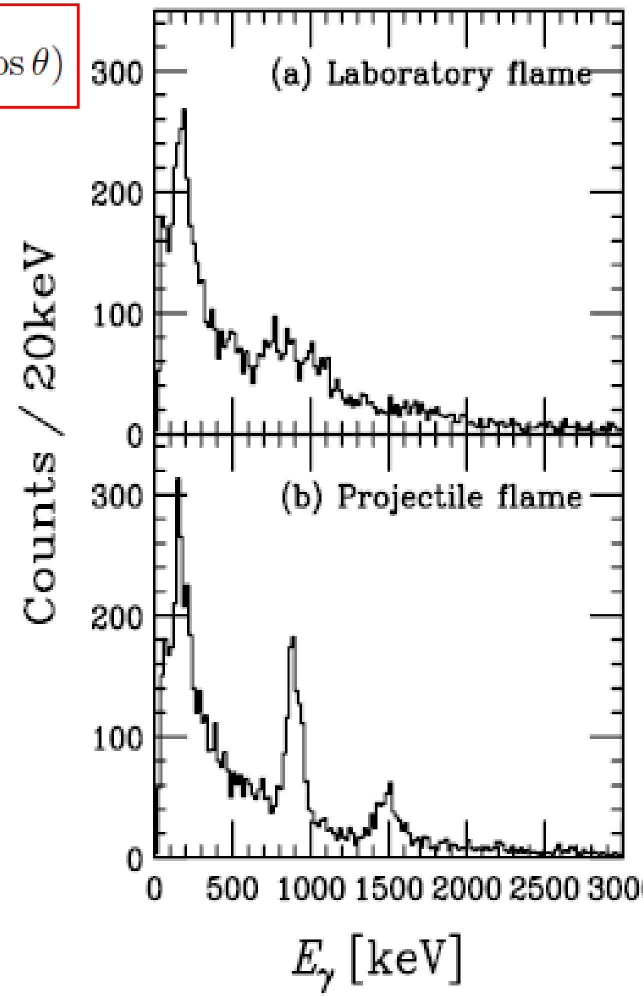
- g-ray source is moving with $\beta \sim 0.3$
  - •Doppler shifted → need to be corrected for

$$E_\gamma^{\rm proj} = E_\gamma^{\rm lab}\gamma(1 - \beta\cos\theta)$$



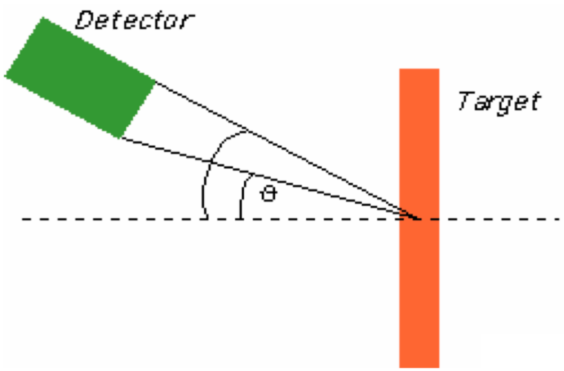


^{32}Mg(p,p') β~0.3

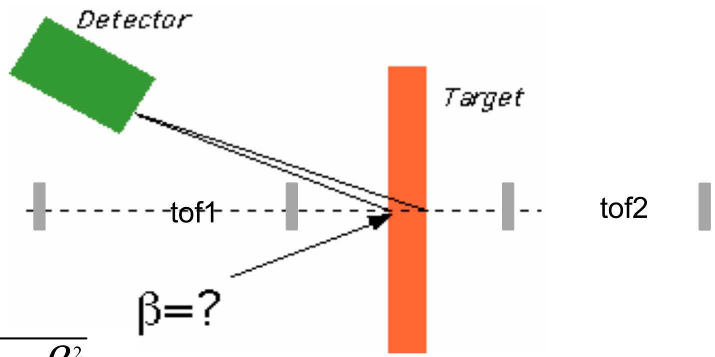


H.Hasegawa, Master's thesis, Rikkyo Univ., 2003

Doppler broadening due to finite opening angle of detector



Doppler broadening due to slowing down of projectile in target



$$E_\gamma = E_\gamma^{proj} \frac{\sqrt{1-\beta^2}}{1-\beta\cos\theta}$$

$$\left(\frac{\Delta E_\gamma}{E_\gamma}\right)^2 = \left(\frac{\beta\sin\theta}{1-\beta\sin\theta}\right)^2(\Delta\theta)^2 + \left(\frac{\cos\theta-\beta}{(1-\beta^2)(1-\beta\cos\theta)}\right)^2(\Delta\beta)^2 + \left(\frac{\Delta E_{intr}}{E_\gamma}\right)^2$$

- $\Delta\theta$ is determined by detector's ability to reconstruct first γ-ray interaction point
- $\Delta\beta$ is determined by target thickness
- ΔE is determined by detector
- Old new paradigm for fast beam experiments with non-4π γ-ray detectors:
  Experimenter trades energy resolution ($\Delta\theta$) versus efficiency

Energy resolution in γ ray measurement ---- design of the array

Dependent on angular resolution / target thickness / detector resolution

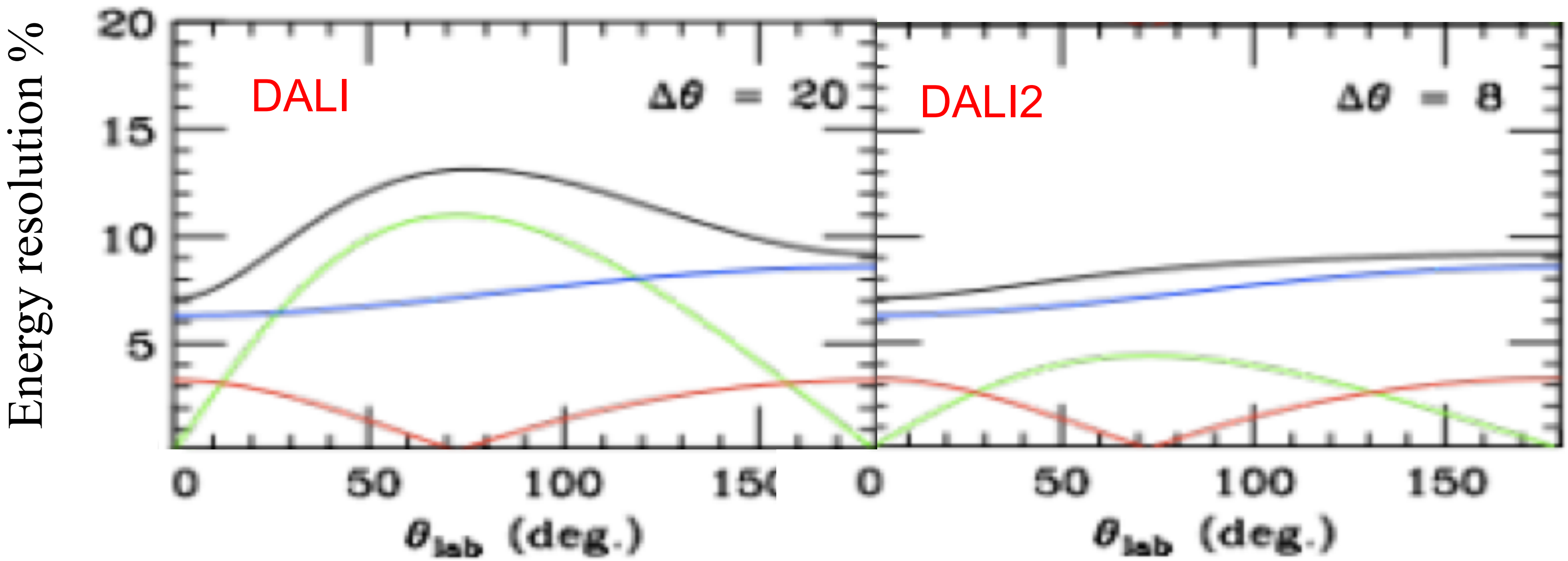
$$E_{\tilde{a}}^{\mathrm{proj}} = \gamma\left(1-\beta\cos\theta^{\mathrm{lab}}\right)E_{\tilde{a}}^{\mathrm{lab}}$$

$$\left(\frac{\Delta E_{\tilde{a}}^{\mathrm{proj}}}{E_{\tilde{a}}^{\mathrm{proj}}}\right)^2 = \left(\frac{\beta\sin\theta^{\mathrm{lab}}}{1-\beta\cos\theta^{\mathrm{lab}}}\right)^2\left(\Delta\theta^{\mathrm{lab}}\right)^2 + \left(\frac{\beta\gamma^2\left(\beta-\cos\theta^{\mathrm{lab}}\right)}{1-\beta\cos\theta^{\mathrm{lab}}}\right)^2\left(\frac{\Delta\beta}{\beta}\right)^2 + \left(\frac{\Delta E_{\tilde{a}}^{\mathrm{lab}}}{E_{\tilde{a}}^{\mathrm{lab}}}\right)^2$$

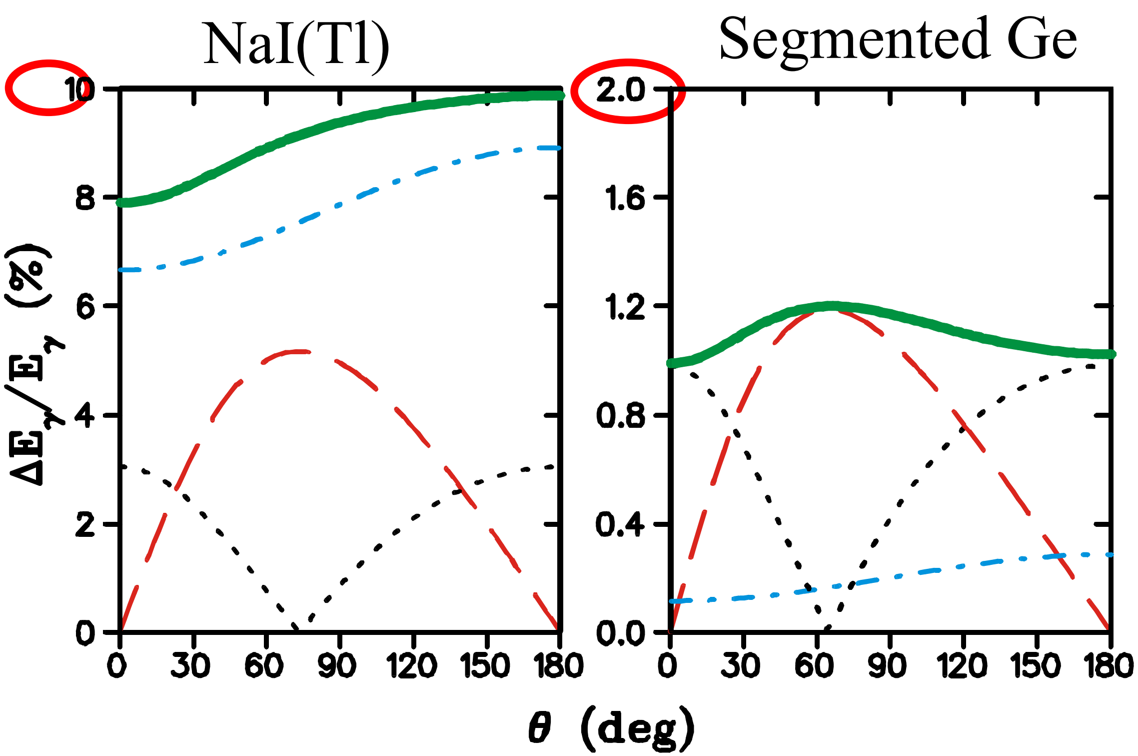
Detector arrangement or Angular resolution

Beam velocity or Target thickness

Intrinsic resolution



Segmented Ge Detectors vs NaI(Tl)



Resolution comparison:
Total
ΔE in target
Opening angle
Intrinsic

Final energy resolution is of the order of 1% with target of order few 100 mg/cm^2
→ detector should have similar or better resolution
→ Energy resolution of ~1% or better
→ Angular resolution of ~ 10 mrad

Parameter estimation

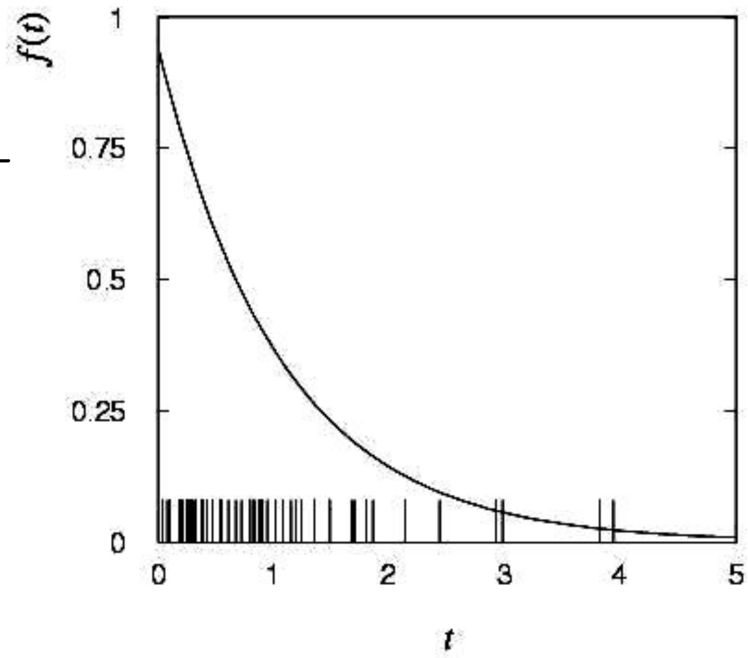
A very common task is to determine the underlying distribution for a measurement. ie. find one (or more) parameters of a pdf f(*x;*a) from a set of measurements $\{x_1, x_2, \ldots \ldots, x_n\}$ -> "estimation"

Example: Radioactive decay

Exponential pdf, $$f(t;\tau) = \frac{1}{\tau} e^{-t/\tau}$$

Experimental data, $t_1, \ldots, t_n$

Task: *determine* τ



Most commonly used methods are:

- Maximum Likelihood Method(MLM)
- Least Squares Method (LSM)

Maximum Likelihood Method

MLH: a general method for estimating parameters of interest from data

- Statement of the maximum likelihood method
  - we have made N measurements of x $\{x_1, x_2, \ldots x_n\}$.
  - we know the probability distribution function that describe x: $f(x;\theta)$.
  - we want to determine the parameter θ.

- How do we use
  - The probability of measuring x_1 is $f(x_1;\theta)dx$.
  - The probability of measuring x_2 is $f(x_2;\theta)dx$.
  - The probability of measuring x_n is $f(x_n;\theta)dx$.

- If the measurements are independent, the probability of getting the measurements is

$$L(\theta) = f(x_1;\theta)\,dx \cdot f(x_2;\theta)\,dx \cdots f(x_n;\theta)\,dx = f(x_1;\theta) \cdot f(x_2;\theta) \cdots f(x_n;\theta)\,dx^n$$

We can drop the dx^n term as it is only a proportionality constant

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

Likelihood function

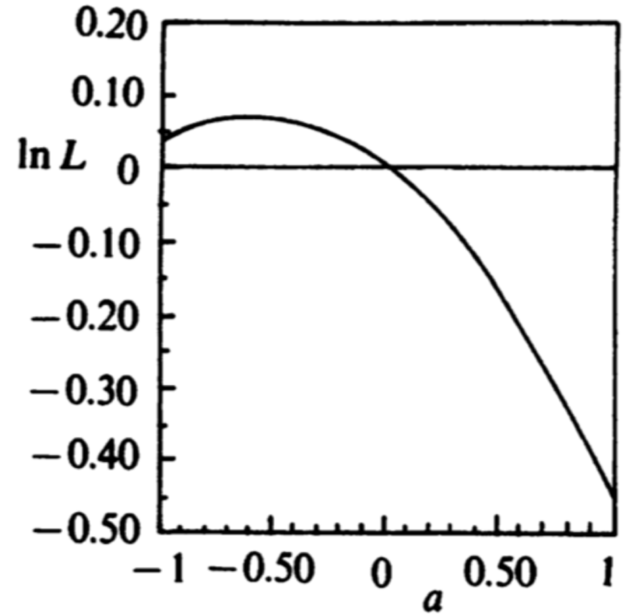
If hypothesis $f(x,\theta)$ and parameter are correct, then we expect a high probability for these measured data sets.

- pick the θ that maximizes L:

$$\left.\frac{\partial L}{\partial \boldsymbol{\theta}}\right|_{\theta=\theta^*} = 0$$

- Both L and lnL have maximum at the same location.
  - maximize lnL rather than L itshelf because lnL converts the product into a summation.

$$\ln L = \sum_{i=1}^{n} \ln f\left(x_i, \theta\right)$$



  - new maximization condition:

$$\left.\frac{\partial \ln L}{\partial \theta}\right|_{\theta=\theta^*} = \sum_{i=1}^{n} \left.\frac{\partial \ln f\left(x_i ; \theta\right)}{\partial \theta}\right|_{\theta=\theta^*} = 0$$

- θ could be an array of parameters (e.g. slope and intercept) or just a single variable.
- equations to determine θ range from simple linear equations to coupled non-linear equations

Error on Estimate

- Taylor expand ln L(θ) around $\theta = \theta^*$

$$\ln L(\theta) = \ln L(\theta^*) + \frac{\partial \ln L}{\partial \theta}\bigg|_{\theta=\theta^*} (\theta - \theta^*) + \frac{1}{2}\frac{\partial^2 \ln L}{\partial \theta^2}\bigg|_{\theta=\theta^*} (\theta - \theta^*)^2 + \ldots$$

$$\ln L(\theta) \approx \ln L(\theta^*) + \frac{1}{2}\frac{\partial^2 \ln L}{\partial \theta^2}\bigg|_{\theta=\theta^*} (\theta - \theta^*)^2$$

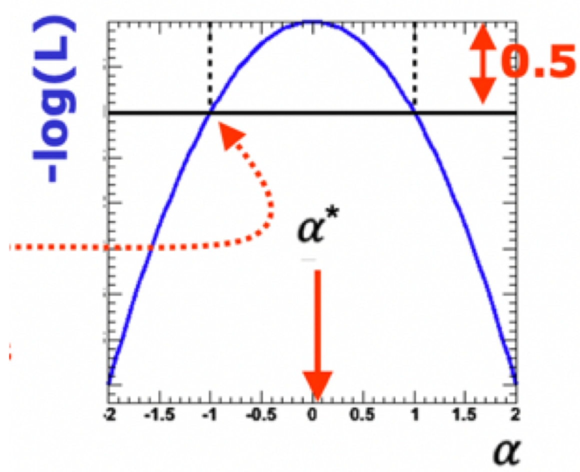
$$L(\theta) \approx const * e^{\frac{1}{2}\frac{\partial^2 lnL}{\partial\theta^2}|_{theta=\theta^*}(\theta-\theta^*)^2}$$

L(α) is Gaussian distributed!

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

$$\sigma(\theta)^2 = -\left(\frac{\partial^2 \ln L}{\partial \theta^2}\bigg|_{\theta=\theta^*}\right)^{-1}$$

$$\ln L(\theta) \approx \ln L(\theta^*) + \frac{(\theta - \theta^*)^2}{2\sigma_\theta^2}$$

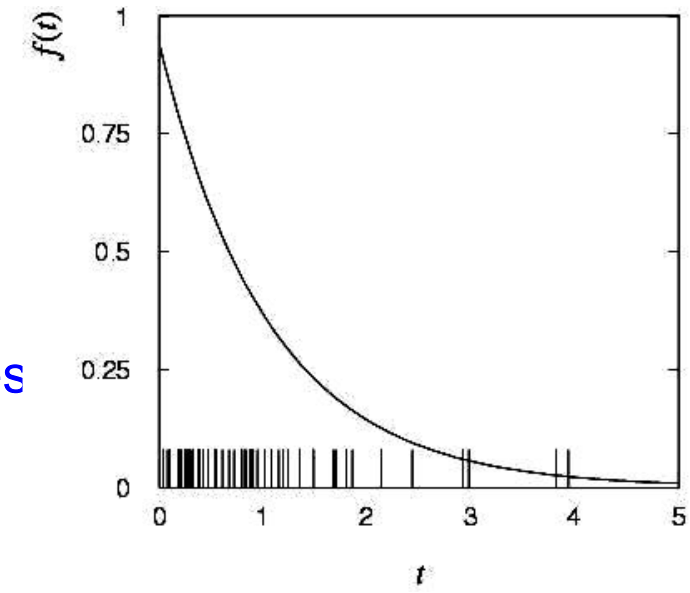


example: parameter of exponential pdf

Consider exponential pdf, $f(t;\tau) = \frac{1}{\tau} e^{-t/\tau}$

Suppose we have data, $t_1, \ldots, t_n$

The likelihood function is $L(\tau) = \prod_{i=1}^{n} \frac{1}{\tau} e^{-t_i/\tau}$



The value of *t* for which *L*(t) is maximum also gives the maximum value of its logarithm :

$$\ln L(\tau) = \sum_{i=1}^{n} \ln f(t_i;\tau) = \sum_{i=1}^{n} \left(\ln \frac{1}{\tau} - \frac{t_i}{\tau} \right)$$

Find its maximum by setting $\frac{\partial \ln L(\tau)}{\partial \tau} = 0$, $\rightarrow$ $\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} t_i$ $\quad \sigma_\tau^2 = \frac{\hat{\tau}^2}{n}$

Monte Carlo test:

generate 50 values
using *t* = 1:

We find the ML estimate: $\hat{\tau} = 1.062$

$\hat{\sigma}_{\hat{\tau}} = 0.150$

- Assume we can measure all times up to limit T
- $f(t;\tau)$ need to be renormalized:

$$f(t;\tau) = \frac{1}{\tau}\frac{e^{-t/\tau}}{(1-e^{-T/\tau})}$$

$$\ln L = \sum_i \left(-\ln\tau - \frac{t_i}{\tau} - \ln\left(1 - e^{-T/\tau}\right)\right)$$

$$\delta \ln L/\delta\tau = 0 \quad \rightarrow \quad \hat{\tau} = \frac{1}{N}\sum t_i + \frac{1}{N}\sum \frac{Te^{-T/\hat{\tau}}}{1-e^{-T/\hat{\tau}}}$$

选读

Extended Maximum Likelihood

- Consider n observations of a random variable *x* distributed according to a p.d.f. $f(x;\theta)$, with unknown parameters $\theta = (\theta_1, \dots \theta_m)$. *Data sample :* $x_1, x_2, \dots, x_n$.
- Often number of observed events n is itself a Poisson random variable with mean value ν.

$$L(\nu, \boldsymbol{\theta}) = \frac{\nu^n}{n!} e^{-\nu} \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}) = \frac{e^{-\nu}}{n!} \prod_{i=1}^{n} \nu f(x_i; \boldsymbol{\theta})$$

This is called extended Likelihood function. Now the sample size *n* defined to be part of the result of the experiment.

$$\ln L(\nu, \theta) = -\nu(\theta) + \sum_i \ln[\nu(\theta) f(x_i, \theta)] + \text{ const}$$

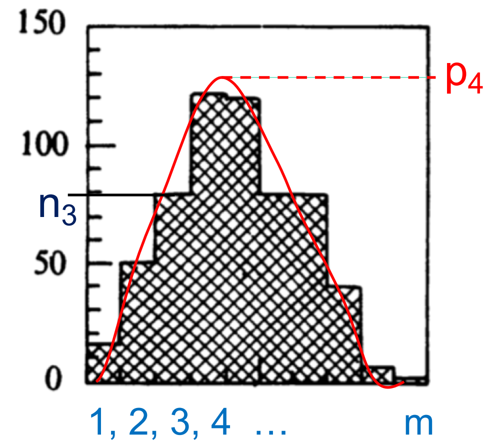
e.g. angles of the outgoing particles, depend on parameters such as particle masses and coupling constants. The number of observed events would fluctuate if one were to repeat the experiment many times, each time with the same integrated luminosity, and not with the same number of events. $\nu = \sigma(m, c) L \varepsilon$

Adding ν as measurement to LH improves resolution on θ (on mass)

If ν is independent of θ, it is the same as normal LH.

Multinomial Distribution

- Generalization of binomial distribution to m possible discrete outcomes for each event:
  - N is total number of trials， probability for "outcome k": p_k
  - Probability to obtaining $(n_1, n_2, \ldots n_m)$ outcomes is given by:



$$f(n_1, \ldots, n_m; N, p_1, \ldots, p_m) = \frac{N!}{n_1! n_2! \ldots n_m!} p_1^{n_1} p_2^{n_2} \ldots p_m^{n_m}$$

Example:

- Throwing a dice 10 times;
  getting 2 * "1" , 1 * "3" , 1 * "4" , 2 * "5" , 4 * "6"

- Consider the three possible outcomes: i, j and everything else.

$$f(n_i, n_j; N, p_i, p_j) = \frac{N!}{n_i! n_j! (N - n_i - n_j)!} p_i^{n_i} p_j^{n_j} (1 - p_i - p_j)^{N - n_i - n_j}$$

Binned Maximum Likelihood (I)

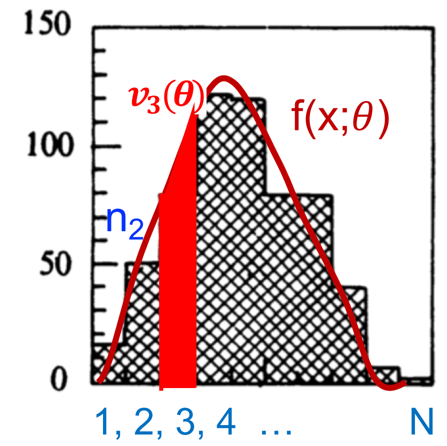
- Consider n_{tot} observations of a random variable *x* distributed according to a p.d.f. $f(x;\theta)$ for which we would like to estimate the unknown parameter $\theta = (\theta_1, \theta_2, \dots \theta_m)$.

- For very large data samples, the log-likelihood function becomes difficult to compute. In such cases, one usually makes a histogram, yielding a certain number of entries $n = (n_1, n_2 ... , n_N)$ in *N* bins.

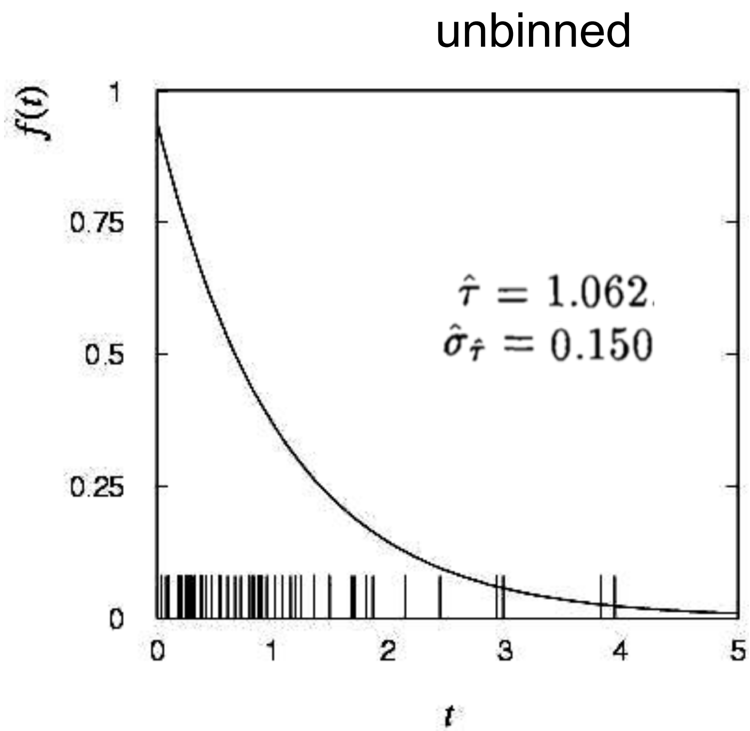
- Compute the number of expected entries in a bin $$\nu_i(\boldsymbol{\theta}) = n_{\text{tot}} \int_{x_i^{\min}}^{x_i^{\max}} f(x;\boldsymbol{\theta})dx$$

$$f_{\text{joint}}\ (\mathbf{n};\boldsymbol{\nu}) = \frac{n_{\text{tot}}\ !}{n_1!\dots n_N!}\left(\frac{\nu_1}{n_{\text{tot}}}\right)^{n_1}\cdots\left(\frac{\nu_N}{n_{\text{tot}}}\right)^{n_N}$$

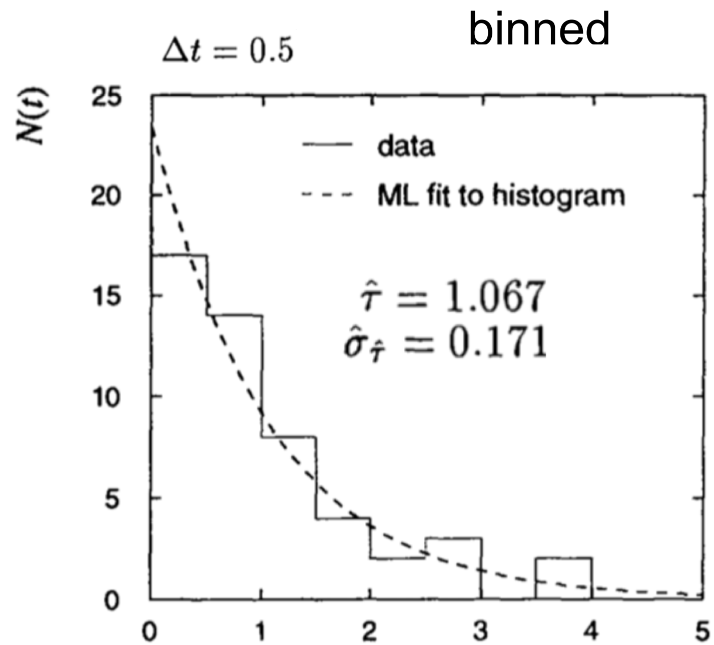
$$\ln(L(\theta)) = \sum_{i=1}^{N} n_i \ln \nu_i(\theta) + \text{ const}$$



- Uncertainties are slightly larger than in unbinned fit
- limit of very small bins



$\Longrightarrow$



选读

Binned Maximum Likelihood (II)

- One may regard the total number of entries as random variable from a Poisson distribution with mean ν_{tot}.

$$f_{\text{joint}}\ (\mathbf{n};\nu) = \frac{\nu_{\text{tot}}^{n_{\text{tot}}}\ e^{-\nu_{\text{tot}}}}{n_{\text{tot}}\ !} \frac{n_{\text{tot}}\ !}{n_1!\dots n_N!} \left(\frac{\nu_1}{\nu_{\text{tot}}}\right)^{n_1} \cdots \left(\frac{\nu_N}{\nu_{\text{tot}}}\right)^{n_N}$$

Where $\nu_{\text{tot}} = \sum_{i=1}^{N} \nu_i$ and $n_{\text{tot}} = \sum_{i=1}^{N} n_i$

$$f_{\text{joint}}(\mathbf{n};\nu) = \prod_{i=1}^{N} \frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i} \qquad \nu_i(\nu_{\text{tot}},\boldsymbol{\theta}) = \nu_{\text{tot}} \int_{x_1^{\min}}^{x_i^{\max}} f(x;\boldsymbol{\theta})dx$$

equivalent to treating the number of entries in each bin as an independent Poisson random variable n_i with mean value ν_i.

$$\ln L(\nu_{tot},\theta) = -\nu_{tot} + \sum_{i=1}^{N} n_i \ln \nu_i(\nu_{tot},\theta)$$

- This is extended LH for binned case. If there is any relation between uncertainties on get smaller, otherwise stay the same.

The method of least squares

- Measurements y_i (eg. differential cross section) with errors σ_i at lots of known points x_i
- A theory gives y=f(x; θ) depending on (unknown) parameter θ
- Want to extract a from the data.
- If errors on data points Gaussian:

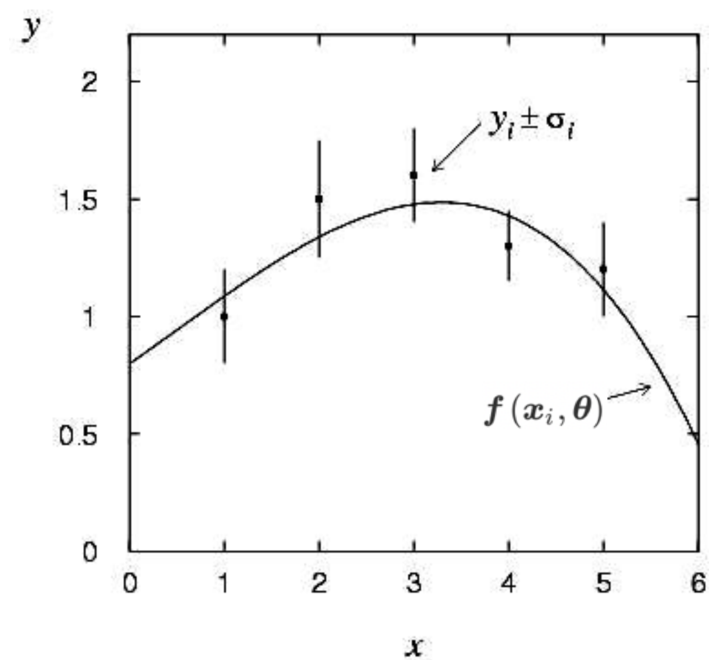
The probability of a particular y_i, for a given x_i is

$$P(y_i;\theta) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-[y_i - f(x_i;\theta)]^2/2\sigma_i^2}$$

$$\ln L(\theta) = -\frac{1}{2}\sum\left[\frac{y_i - f(x_i;\theta)}{\sigma_i}\right]^2 - \sum \ln \sigma_i\sqrt{2\pi}$$

Maximize $\ln L(\theta)$ means minimizing

$$\chi^2 = \sum_i \left[\frac{y_i - f(x_i,\theta)}{\sigma_i}\right]^2$$



Often minimize χ^2 numerically (e.g. program **MINUIT**).

Some Remarks on χ^2

- By definion of $\chi^2 = \sum_i \frac{\left(y_i - f(x_i)\right)^2}{\sigma_i^2}$ expect ~1 per data point.
- More precisely, expect χ^2 ~1 per number of degree of freedom (ndf)

$$N_{ndf} = N_{\text{data points}} - N_{\text{fit parameters}}$$

e.g. if we fitted a Gaussian, there were 3 parameters

- χ^2 / ndf provides a figure of merit for how well theory describes data

Simplest Example: Straight Line Fit

For simplicity, suppose line must go through origin:

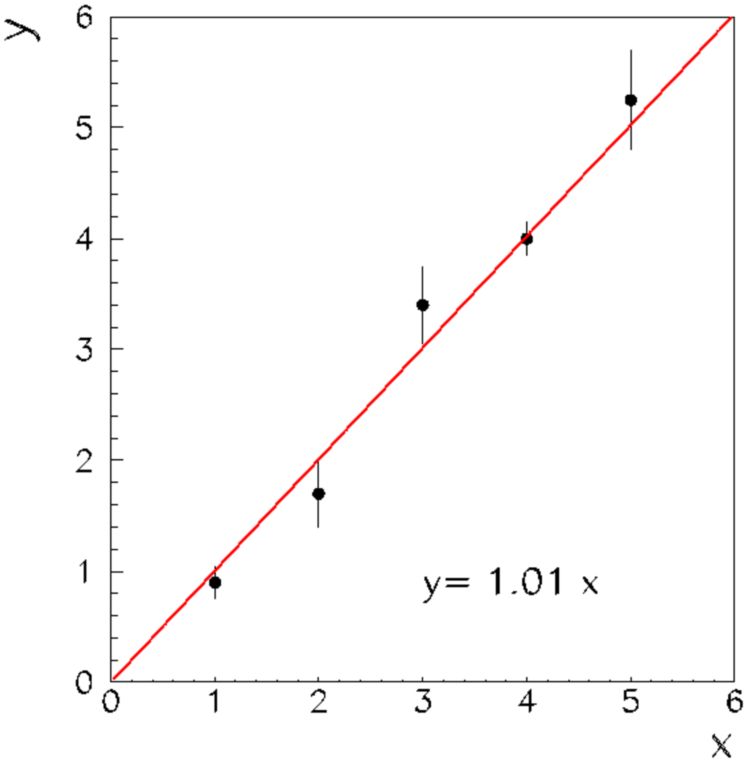
y=f(x)=mx

$$\chi^2 = \sum_i \frac{(y_i - mx_i)^2}{\sigma_i^2}$$

Minimise with respect to m….

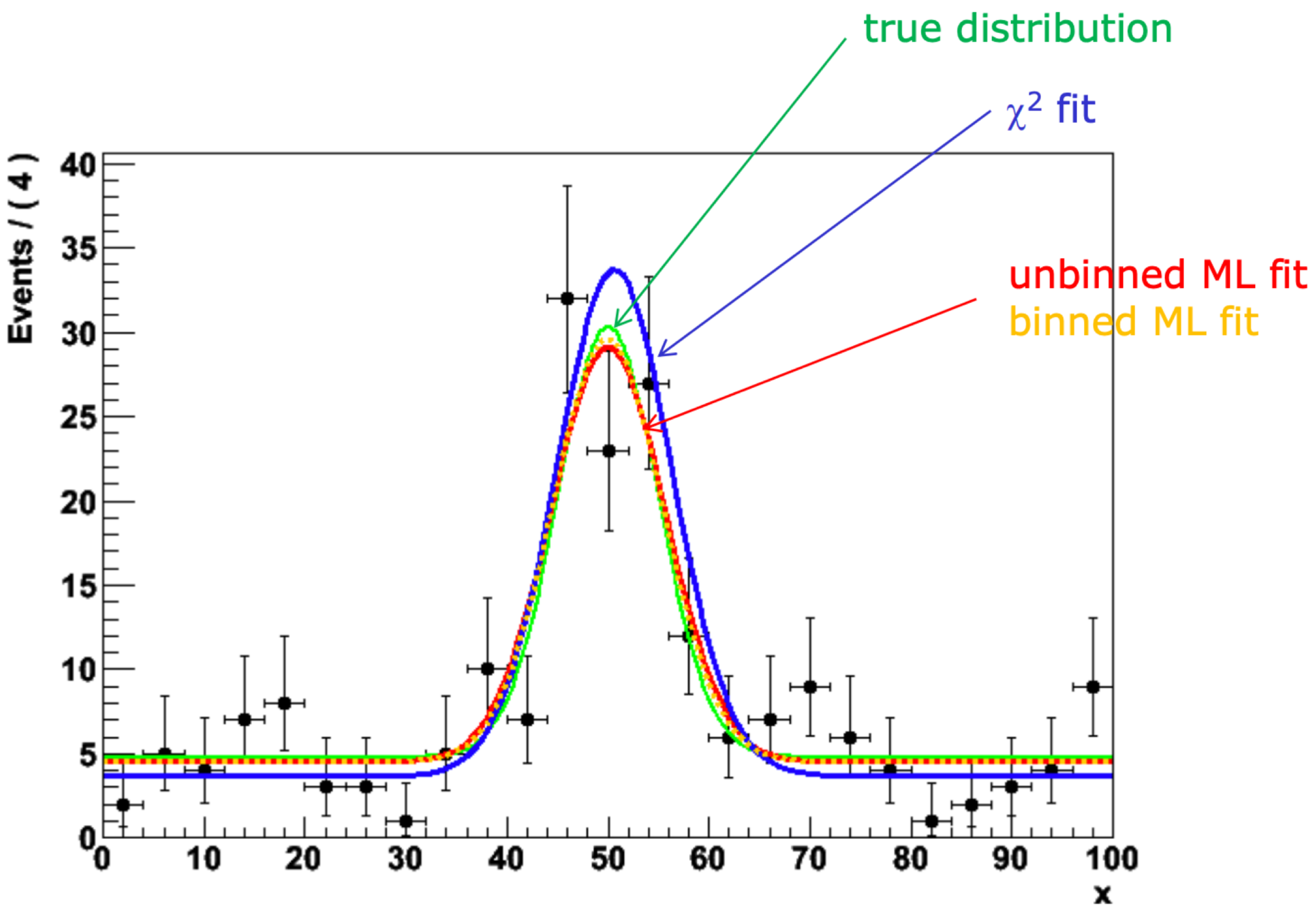
$$\frac{d\chi^2}{dm} = \sum_i -2x_i \frac{(y_i - mx_i)}{\sigma_i^2} = 0$$

$$m = \frac{\sum_i x_i y_i / \sigma_i^2}{\sum_i x_i^2 / \sigma_i^2}$$



Example of χ^2 vs ML fit

- Example with many low statistics bins

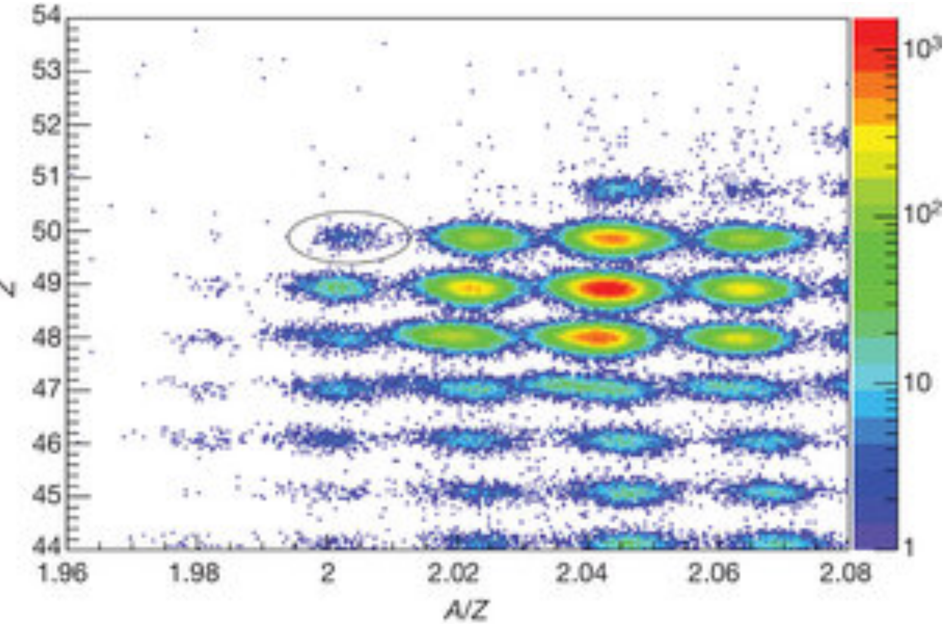


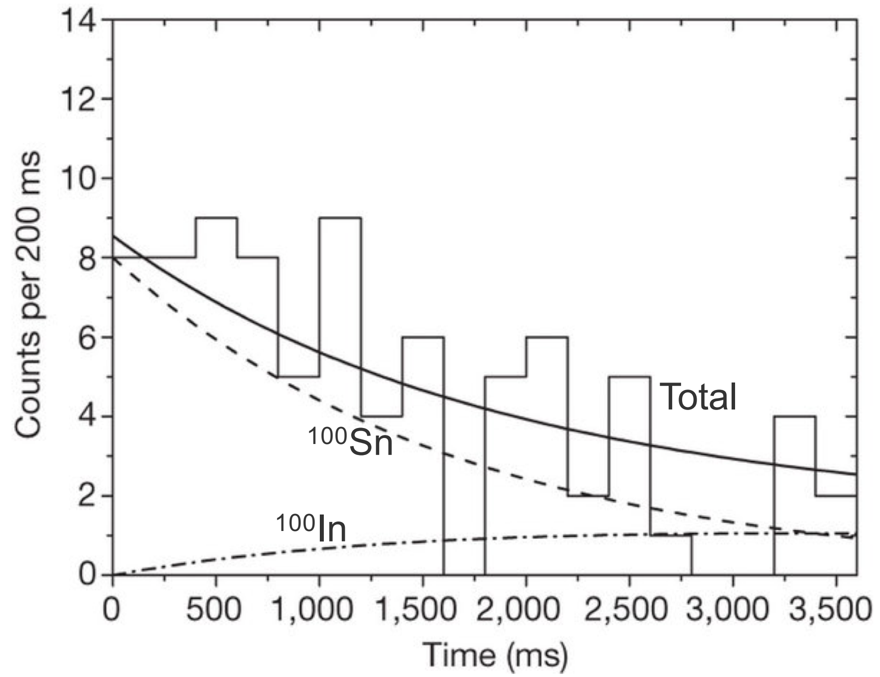
Maximum Likelihood or χ^2 – What should you use?

- χ^2 fit is fastest, easiest
  - Works fine at high statistics
  - Gives absolute goodness-of-fit indication
  - Make (incorrect) Gaussian error assumption on low statistics bins
  - Has bias proportional to 1/N
  - Misses information with feature size < bin size

- Full Maximum Likelihood estimators most robust
  - No Gaussian assumption made at low statistics
  - No information lost due to binning
  - Gives best error of all methods (especially at low statistics)
  - No intrinsic goodness-of-fit measure, i.e. no way to tell if ‘best’ is actually ‘pretty bad’
  - Has bias proportional to 1/N
  - Can be computationally expensive for large N

C.B.Hinke et al., Nature 486, 341–345 (2012)





In total, 259 ^{100}Sn nuclei (those indicated in the figure) were unambiguously identified.

Time distribution of first decay events

A **maximum-likelihood analysis** with a maximum of three decay events during the correlation time was used to analyse these decay chains. The half-life of 100Sn was deduced to be 1.16 6 0.20 s in the MLH analysis using established values for the lifetimes of the daughter nuclei.

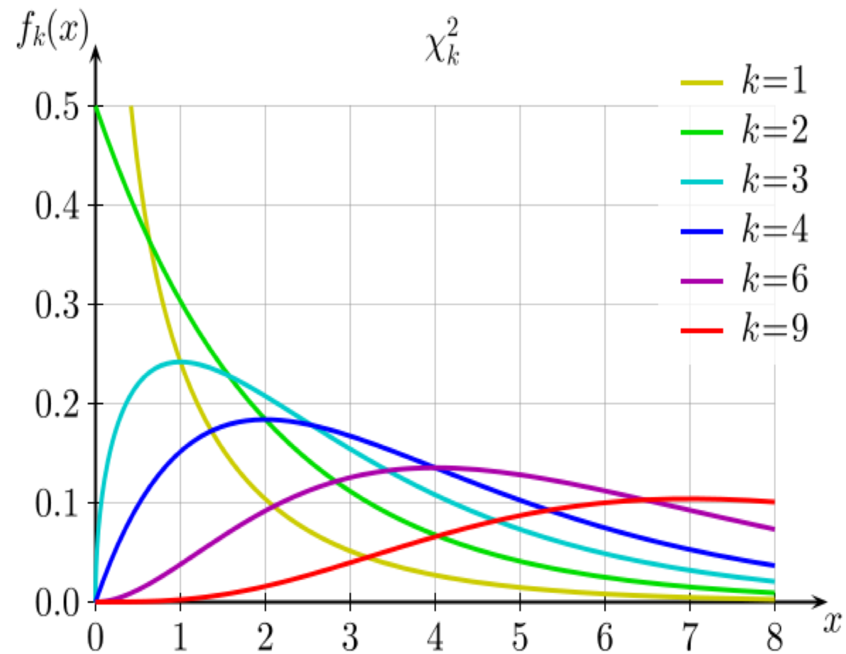
The Chi-Square(χ^2) Distribution

- Important in connection with least-square method.
- If x_1, x_2, … x_n are independent, Gaussian distributed variables, with mean μ and variance σ, then $\chi^2 = \sum[(x_i - \mu_i)/\sigma_i^2]$ is distributed according to χ^2-distribution.

$$f(k,\chi^2) = \frac{(\chi^2/2)^{k/2-1}e^{-\chi^2/2}}{2\Gamma(k/2)}$$

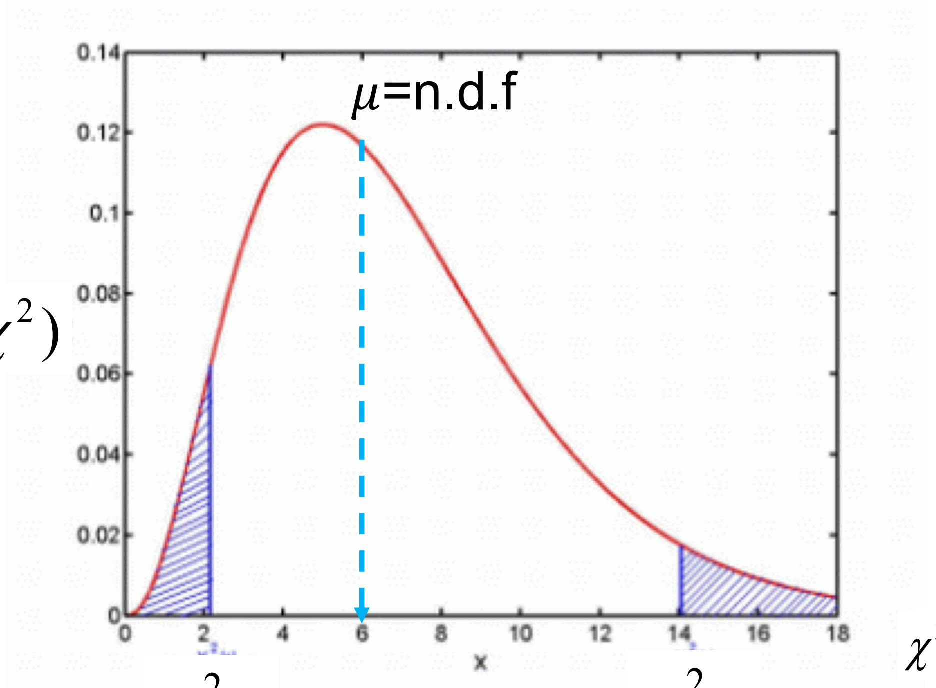
k: number of degrees of freedom
ndf/df/dof

$\Gamma(k/2)$: gamma function



χ^2 -distribution plays an important role in the comparison of measurements with theoretical distributions.

Chi-Square(χ^2) Test - Goodness of Data



$P(\chi^2)$, χ^2, $\chi^2_{0.95}$, $\chi^2_{0.05}$

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2$$

- μ = *k*
- σ² = 2 * *k*

$$P(\chi^2 \geq \chi^2_\alpha) = \int_{\chi^2_\alpha}^{\infty} P(\chi^2)\, d\chi^2 = \alpha$$

Very low probabilities (say less than 0.05) indicate abnormal large fluctuations in data, whereas very high probabilities(greater than 0.95) indicate abnormally small fluctuations.

A perfect fit to the distribution for large samples would yield a probability of 0.5

The chi-square pdf has an expectation value equal to the number of degrees of freedom, so if $\boldsymbol{\chi^2/ndf \approx 1}$, the fit is 'good'.

For example, if for identical, consecutive measurements one gets the following counts in a scaler:

242, 241, 249, 246, 236, 250

$$\overline{N} = \frac{1}{k}\sum_{i=1}^{k} N_i = \frac{1}{6}(242+241+\cdots+250) = \frac{1}{6}\times 1464 = 244$$

$$\chi^2 = \sum_{i=1}^{k}\frac{(N_i-\overline{N})^2}{\overline{N}} = \frac{1}{244}\left[(242-244)^2+\cdots+(250-244)^2\right] = \frac{142}{244} = 0.58$$

Degrees of freedom 6-1=5

Given X^2= 0.58 and d= 5

Calculate

The chance probability, Q, is: 0.9889

https://www.fourmilab.ch/rpkp/experiments/analysis/chiCalc.html

The data are clustered around the mean much closer than one would expect, suspicious !

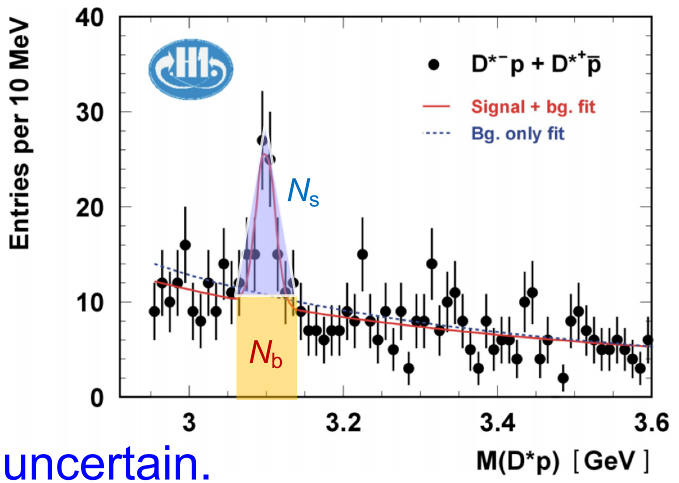
The significance of an observed signal

Suppose we observe *n* events; these can consist of:

N_b events from known processes (background)
N_s events from a new process (signal)

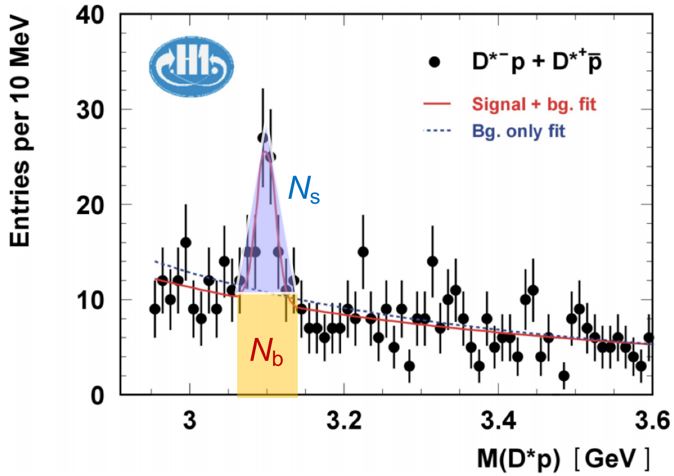
If N_s, N_b are Poisson random variables with means μ_s, μ_b, then $N = N_s + N_b$ is also Poisson, mean = $\mu_s + \mu_b$:

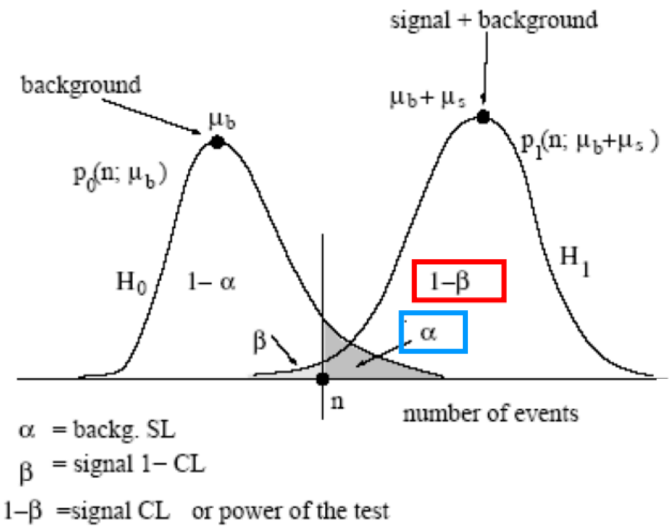
$$P(N; \mu_s, \mu_b) = \frac{(\mu_s + \mu_b)^N}{N!} e^{-(\mu_s + \mu_b)}$$



Sometimes *b* known, other times it is in some way uncertain.

Goals:
(*i*) convince people that $\mu_s \neq 0$ (discovery);
(*ii*) measure or place limits on μ_s, taking into consideration the uncertainty in μ_b.





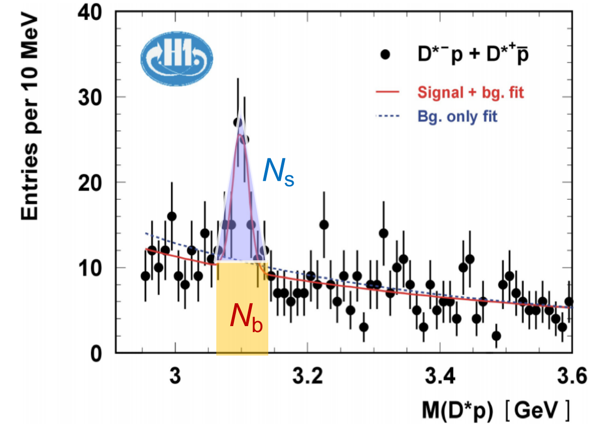
$$P(N;\mu_s,\mu_b) = \frac{(\mu_s+\mu_b)^N}{N!} e^{-(\mu_s+\mu_b)}$$

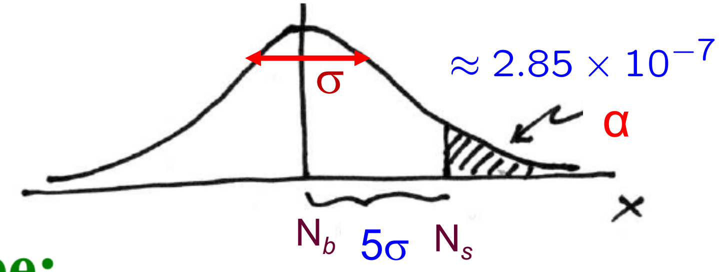
Suppose N_b = 0.5, and we observe N = 5.
Should we claim evidence for a new discovery?
Give α-value for hypothesis s = 0:

$$\alpha\text{ -value} = P(N \geq 5; b = 0.5, s = 0)$$
$$= 1.7 \times 10^{-4} \neq P(s = 0)!$$

Significance from α-value

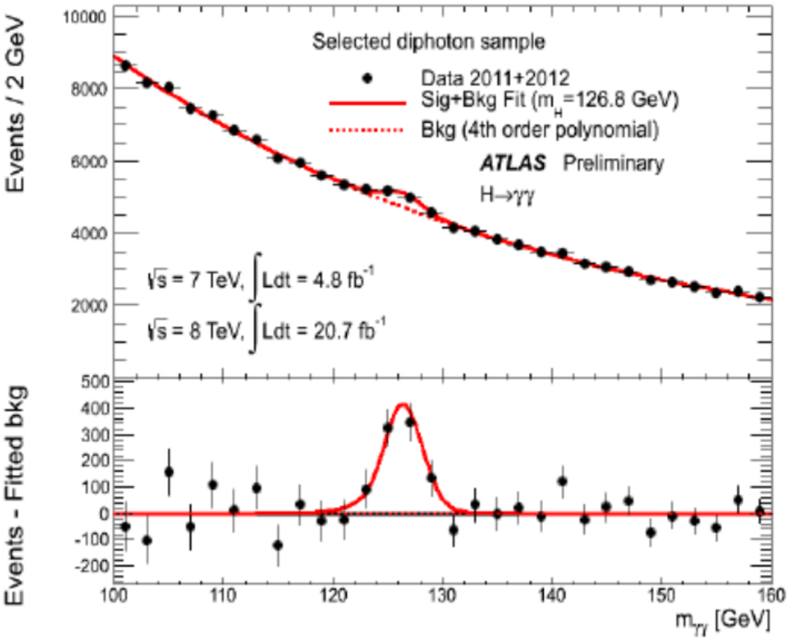
- **In a given amount of data we expect:**
  - **N_B background events**
  - **Statistical error on background $\approx \sqrt{N_B}$**
  - **Systematic error on background = σ_{sys}**
  - **Add errors in quadrature to get σ_{TOT}**
- **Observe $N(>N_B)$ events in data. Could be:**
  - **random fluctuation in $N_B \pm \sigma_{TOT}$ background events**
  - **N_B background events & N_S signal events**
- **Significance $S = N_S/\sigma_{TOT}$**





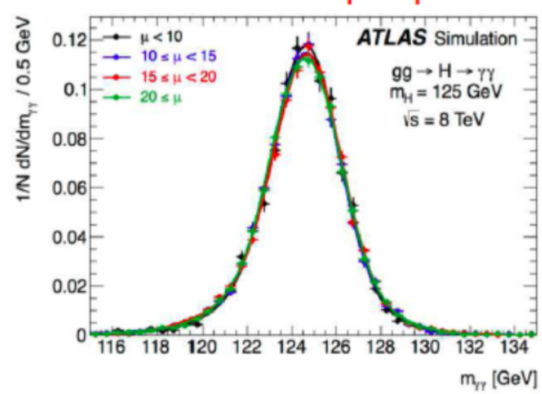
α	$\leq 2.8 \cdot 10^{-7}$	5σ	discovery
α	$\leq 1.3 \cdot 10^{-3}$	3σ	strong evidence
α	$\leq 2.3 \cdot 10^{-2}$	2σ	weak evidence

Discovery channel H -> γ γ



H -> γ + γ

Mass resolution is pile-up robust



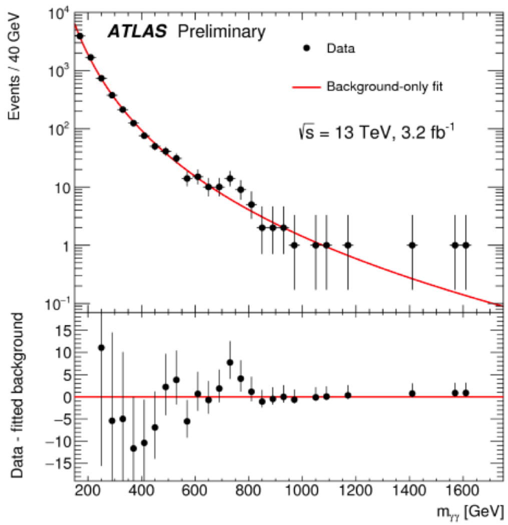
Simple topology: two high-E_T (>40,30 GeV) isolated photons

- Observed local significance of the excess: **7.4σ** (4.1σ expected for SM Higgs)
- Best mass fit: **126.8 ± 0.2 (stat) ± 0.7 (syst) GeV** → Systematics fully dominated by γ-energy scale
- Best fit of signal strength @ this mass **$\mu=1.65^{+0.34}_{-0.30}$ = 1.65 ± 0.24 (stat) $^{+0.25}_{-0.18}$ (syst)**
  [consistent across various categories]

LHC's 750 GeV bump

- 2015 data ($\sqrt{s}$=13 TeV, 3.2 fb $^{-1}$) had an excess at $m_{\gamma\gamma}{\sim}750\ \mathrm{GeV}/c^2$.

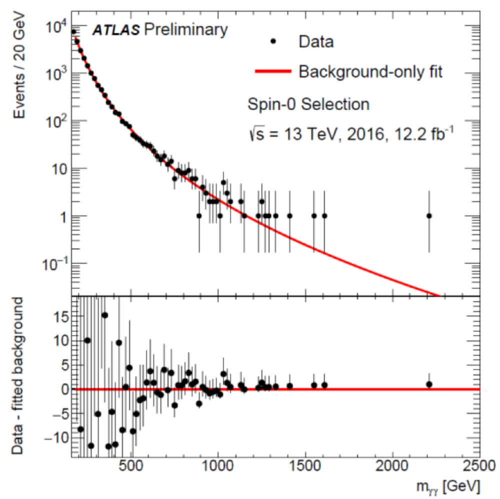
2015 data 3.2 fb $^{-1}$



$\sim 3.8\sigma$

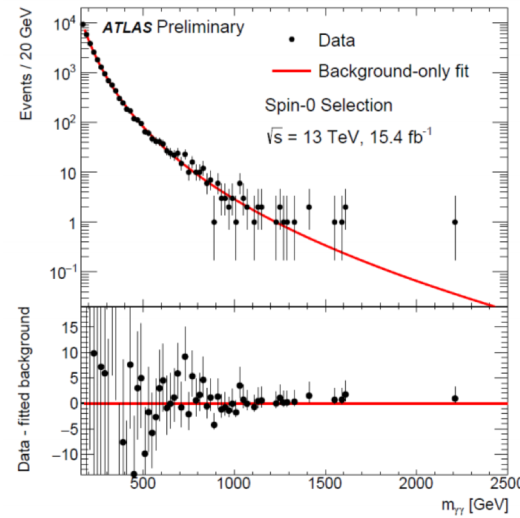
- CMS had similar excess with local $\sim 3\sigma$

ATLAS 2016 data: 12.2 fb^{-1}



2016 data in agreement with expected background within 1σ.

ATLAS 2015+2016 data: 15.4 fb^{-1}



With 2015+2016 combined data, largest local significance in 700-800 GeV was 2.3σ for 710 GeV.