International Atomic Energy Agency

Basic Concepts

Random variable, probability distribution, mean, variance, standard variation, uncertainty, covariance

Naohiko Otsuka

IAEA Nuclear Data Section

Random Variables

We believe physical quantity has a unique true value (within classical approx.) which is what evaluators want to determine.

Experimental results are stochastic due to statistical fluctuation and limitations of the measurement procedures – *random variables*.

Example: Number of events *N*

By repeating a counting experiment n times, we obtain a set of counting number

$\{N\}=N_1, N_2,\ldots,N_n$

They do not agree in general. N is a random variable.

Probability Distribution

We believe that the results distribute with a peak around the true value (probability distribution).

In general, this distribution is described by

P_k : probability for a discrete random variable k.

P(x): probability (density) for a continuous random variable x.

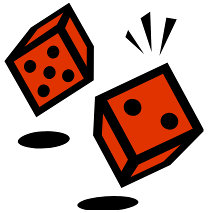
By definition,

$\Sigma_k P_k = 1$ (discrete random variable)

$\int dx\, P(x)=1$ (continuous random variable)

Discrete Random Variables – One Dice

Probability to find a value (1, 2, 3, 4, 5 or 6) on a dice

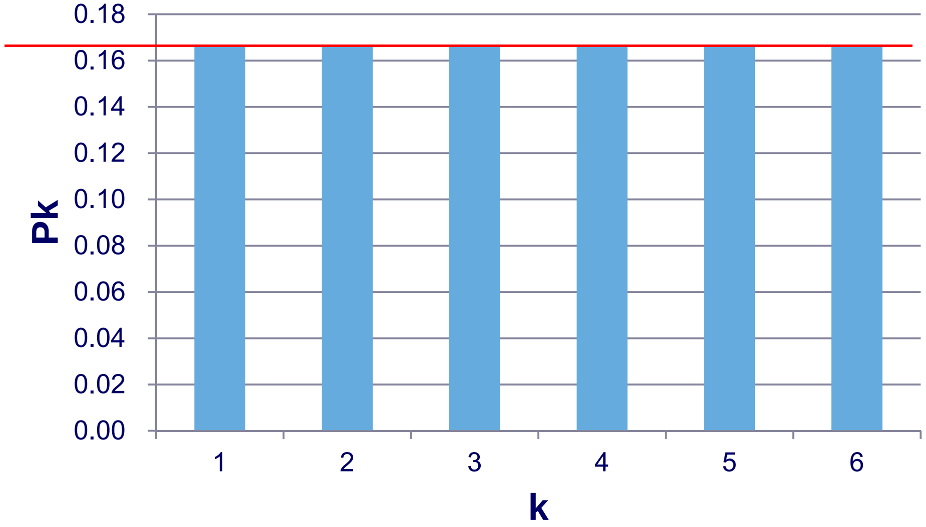


k (=1,2,, or 6): random variable

P_k (= 1/6 for each k): probability distribution

k	1	2	3	4	5	6
P_k	1/6	1/6	1/6	1/6	1/6	1/6

1/6 →



Discrete Random Variables: Sum from Two Dices

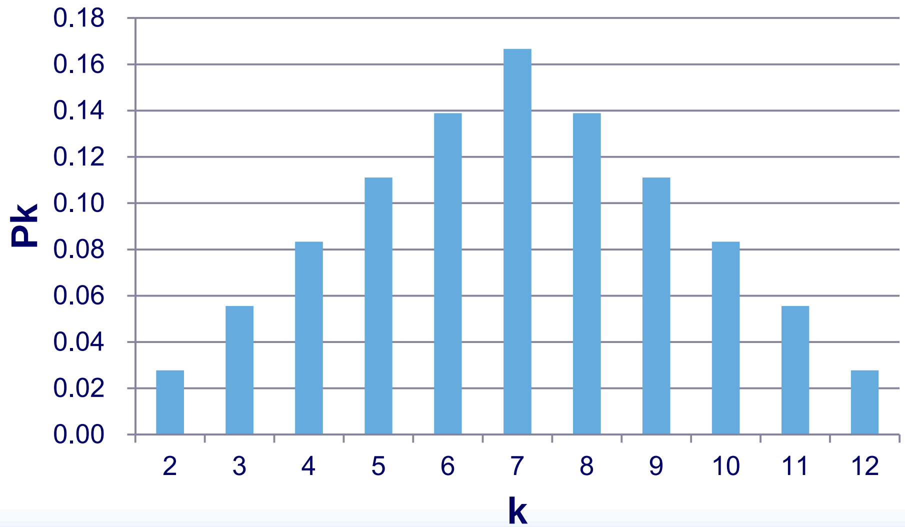
Probability to find k=i+j with i and j (=1,,6) on two dices

i=(1,2,, or 6) – random variable

j=(1,2,, or 6) – random variable

→ k=(2,3,4,…,12) is also a random variables.

k	2	3	4	5	6	7	8	9	10	11	12
P_k	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36



Mean, Variance and Standard Deviation (Discrete Variable)

Definition of mean, variance and standard deviation for a discrete random variable k following the probability distribution P_k:

- **Mean**

  $\langle k \rangle = \Sigma_{k=1,n}\, k \cdot P_k$

- **Variance**

  $v = \langle (k-\langle k \rangle)^2 \rangle = \langle k^2 \rangle - \langle k \rangle^2$

- **Standard deviation**

  $\Delta k = (v)^{1/2}$

Mean and standard deviation are often adopted as "best estimate" and "uncertainty" (will be discussed later).

Mean, Variance and Standard Deviation (Continuous Variable)

For a continuous random variable x, similarly

- **Mean**

  $\langle x \rangle = \int dx\, x \cdot P(x)$

- **Variance**

  $v = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2$

- **Standard deviation**

  $\Delta x = (v)^{1/2}$

Mean, Variance and Standard Deviation (Discrete Variable)

Exercise:

Prove that $v=\langle(k-\langle k\rangle)^2\rangle=\langle k^2\rangle-\langle k\rangle^2$

$v = \langle(k-\langle k\rangle)^2\rangle$

$=\langle k^2-2k\langle k\rangle+\langle k\rangle^2\rangle$

$=\langle k^2\rangle-\langle 2k\langle k\rangle\rangle+\langle\langle k\rangle^2\rangle$

$= \langle k^2\rangle-2\langle k\rangle\langle k\rangle+\langle k\rangle^2$

$= \langle k^2\rangle-\langle k\rangle^2$

$\langle\ldots\langle x\rangle\ldots\rangle = \langle x\rangle\langle\ldots\rangle$

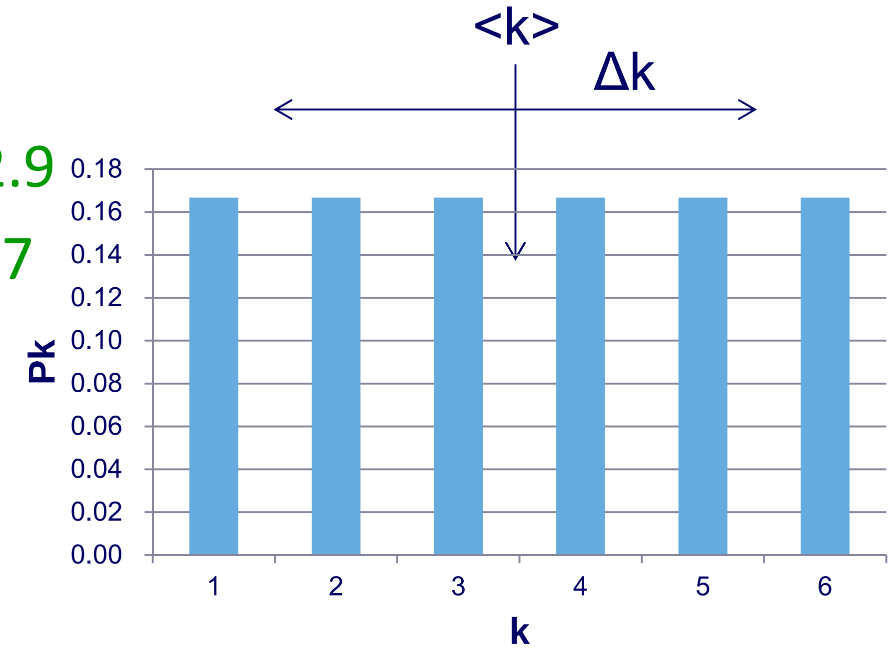
Mean, Variance and Standard Deviation: One Dice

k	1	2	3	4	5	6
P_k	1/6	1/6	1/6	1/6	1/6	1/6

k: number on a dice

- mean $\langle k\rangle = \Sigma_{k=1,6}\, k\cdot P_k \sim 3.5$
- variance $v = \Sigma_{k=1,6}\, k^2\cdot P_k - \langle k\rangle^2 \sim 2.9$
- standard deviation $\Delta k=(v)^{1/2} \sim 1.7$

Even if the probability is equally distributed, we can define mean and standard deviation.



There is *no* concept of “true value” and “uncertainty”.

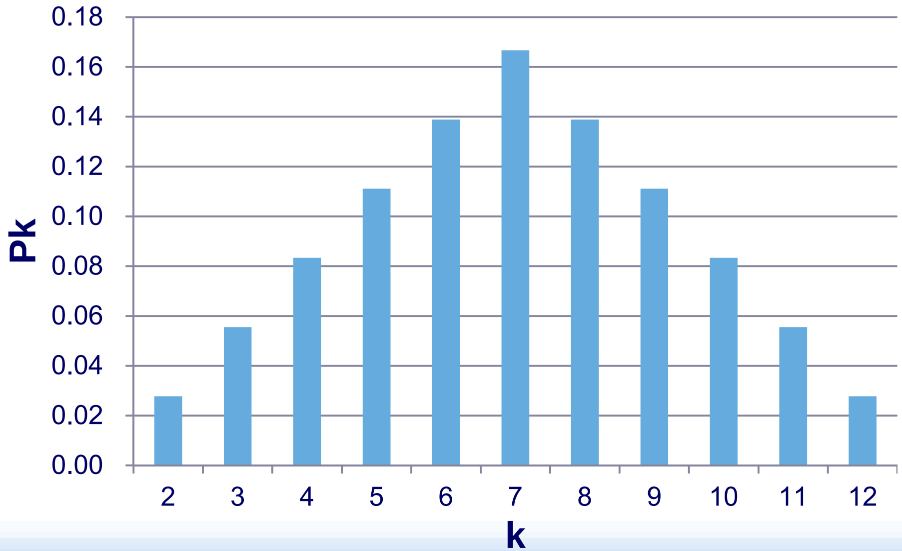
Mean, Variance and Standard Deviation: Two Dices

k=m+n with m and n (m, n=1,,6) on two dices.

k	2	3	4	5	6	7	8	9	10	11	12
P_k	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Calculate mean, variance and standard deviation of k.

- mean $\langle k \rangle = \Sigma_{k=2,12}\ k \cdot P_k$=?
- variance $v = \Sigma_{k=2,12}\ k^2 \cdot P_k - \langle k \rangle^2$=?
- standard deviation $\Delta k = (v)^{1/2}$=?

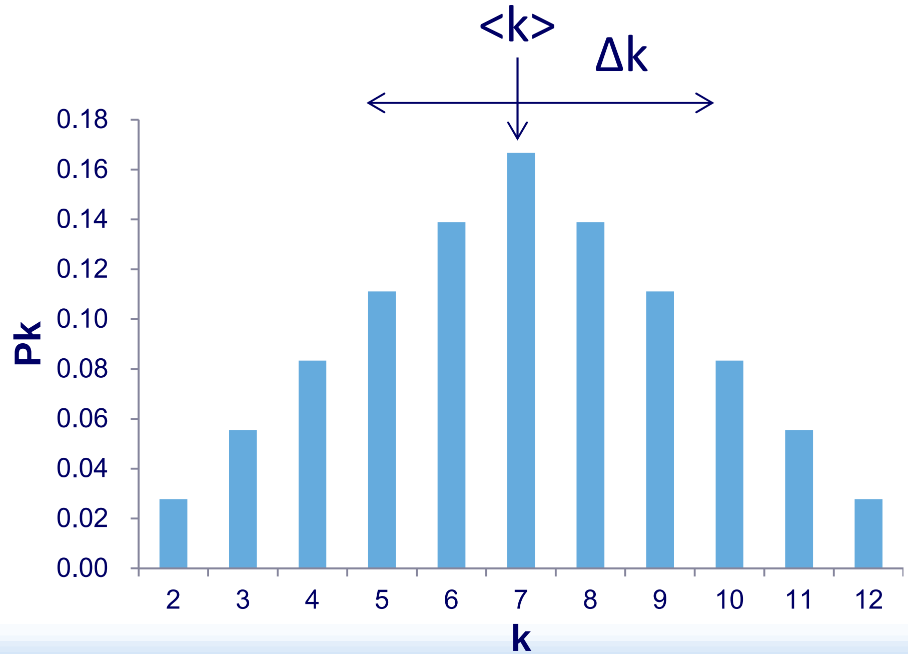


Mean, Variance and Standard Deviation: Two Dices

k=m+n with m and n (m, n=1,,6) on two dices

k	2	3	4	5	6	7	8	9	10	11	12
P_k	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

- mean $\langle k\rangle = \Sigma_k\, k\cdot P_k \sim 7$
- variance $v = \Sigma_k\, k^2\cdot P_k - \langle k\rangle^2 \sim 5.8$
- standard deviation $\Delta k=(v_k)^{1/2} \sim 2.4$



Population and Sample

In general, we cannot know the probability distribution (e.g., 1/6) without experiment.

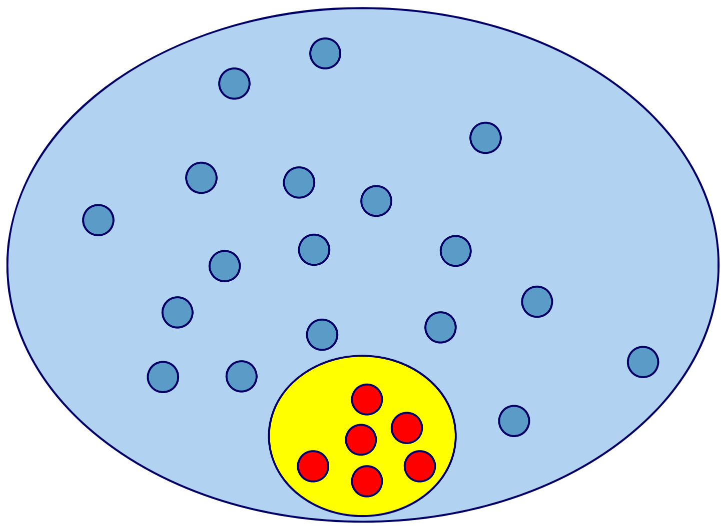
We cannot measure the whole set ("population") to extract the statistical property.

The statistical property of the population may be estimated by sampling (size n).

For a discrete random variable k,

- mean $\langle k \rangle' = \Sigma_{i=1,n}\, k_i / n$
- variance $v' = \Sigma_{i=1,n}\, k_i^2 / n - \langle k \rangle'^2$
- standard deviation $\Delta k' = (v')^{1/2}$

Population (<k>, v, Δk)



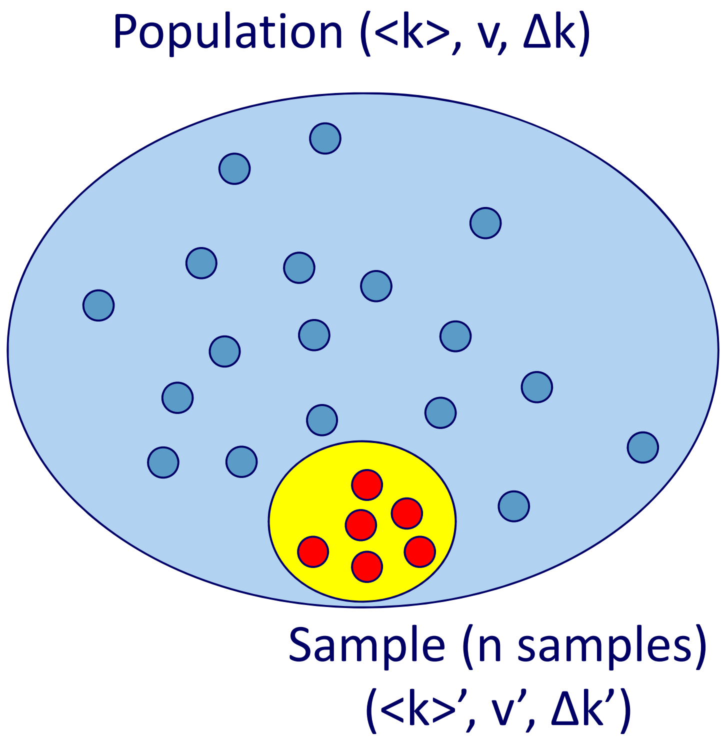
Sample (N samples)
(<k>', v', Δk')

Population and Sample (cont)

The properties of the population is related with those of a size n sample ($\langle k \rangle'$, v', Δk') as follows:

- $\langle k \rangle = \langle k \rangle'$ (equal!)
- $v = v' \, [n/(n-1)]$
- $\Delta k = \Delta k' [(n-1)/2]^{1/2} \, \Gamma[(n-1)/2] / \Gamma(n/2)$

Γ: gamma function
For integer n, $\Gamma(n+1) = n! = n \cdot (n-1) \cdot \ldots \, 2 \cdot 1$.
$\Gamma(n/2) / \Gamma[(n-1)/2] \sim [(n-1.5)/2]^{1/2}$.

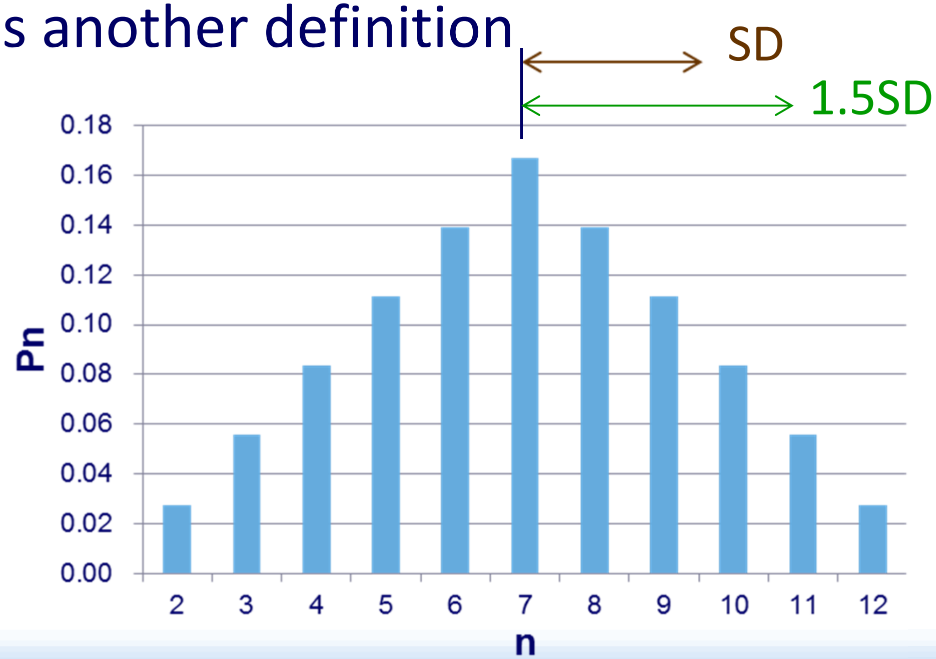


If the sample size n is enough large, n-dependent correction close to 1.

Best Estimate and Uncertainty

Mean and standard deviation (SD) are often adopted as the best estimate and uncertainty, respectively.

The definition of "uncertainty" is *not unique*. For example, one may adopt 1.5SD instead of SD as another definition of the uncertainty.



Unusual Standard Deviation(LEXFOR "Errors")

2. Only uncertainties that are one standard deviation (or the equivalent for systematic uncertainty) are entered in this format. If the author gives 2- or 3-sigma uncertainties, they should be converted to 1-sigma uncertainties before entering. Other types of uncertainty information may be entered in free text.

This asks us to convert uncertainties given in 2SD or 3SD to SD before entering under **DATA-ERR** etc. (SD: standard deviation)

The definition of the uncertainty is usually not seen in literature, and we usually *assume* that 1SD is given, and entre it without conversion.

Example: 1.5 ± 0.3 mb by authors →
Compiler assumes that 0.3 mb is the 1SD (standard deviation).

Definition of Uncertainty
("Review of Particle Physics")

1. PHYSICAL CONSTANTS

Table 1.1. Reviewed 2011 by P.J. Mohr (NIST). Mainly from the "CODATA Recommended Values of the F[illegible] 2010" by P.J. Mohr, B.N. Taylor, and D.B. Newell in `arXiv:1203.5425` and Rev. Mod. Phys. (to be publish[illegible] (beginning with the Fermi coupling constant) comes from the Particle Data Group. The figures in parentheses after the values give the 1-standard-deviation uncertainties in the last digits; the corresponding fractional uncertainties in parts per 10^9 (ppb) are given in the last column. This set of constants (aside from the last group) is recommended for international use by CODATA (the Committee on Data for Science and Technology). The full 2010 CODATA set of constants may be found at `http://physics.nist.gov/constants`. See also P.J. Mohr and D.B. Newell, "Resource Letter FC-1: The Physics of Fundamental Constants," Am. J. Phys, **78** (2010) 338.

Quantity	Symbol, equation	Value	Uncertainty (ppb)
speed of light in vacuum	c	299 792 458 m s^{-1}	exact*
Planck constant	h	6.626 069 57(29)$\times 10^{-34}$ J s	44
Planck constant, reduced	$\hbar \equiv h/2\pi$	1.054 571 726(47)$\times 10^{-34}$ J s	44
		= 6.582 119 28(15)$\times 10^{-22}$ MeV s	22
electron charge magnitude	e	1.602 176 565(35)$\times 10^{-19}$ C = 4.803 204 50(11)$\times 10^{-10}$ esu	22, 22
conversion constant	$\hbar c$	197.326 9718(44) MeV fm	22
conversion constant	$(\hbar c)^2$	0.389 379 338(17) GeV^2 mbarn	44
electron mass	m_e	0.510 998 928(11) MeV/c^2 = 9.109 382 91(40)$\times 10^{-31}$ kg	22, 44

J. Beringer *et al.* (Particle Data Group), Phys. Rev. D**86**, 010001 (2012)

The caption clearly states that 1SD is adopted as the uncertainty in the table.

Example (EXFOR 31686)

```
ERR-ANALYS (ERR-T) 1 sigma converted from 2 sigma by compiler
           (ERR-1) Uncertainty due to spectrometry (5% at 2
                   sigma)
STATUS     (APRVD) Irina Glagolenko (2010-01-14)
HISTORY    (20091117C) On
           (20100114A) On. Approved by authors.
ENDBIB               12
COMMON                1          3
ERR-1
PER-CENT
 2.5
ENDCOMMON             3
ENDSUBENT            19
```

Table 1 Comparison of the U-235 cumulative thermal fission yield data measured by ICP-MS method relative to Y-89 and La-139 with ENDF/B-VII data

Isotope	U-235 cumulative thermal fission yield measured relative to Y-89[a], %
Rb-87	2.65 ± 0.08[b]
Y-89	4.73 ± 0.14
Tc-99	5.99 ± 0.18
Ru-101	5.09 ± 0.15
Ru-102	4.20 ± 0.13
Rh-103	2.92 ± 0.09
Ru-104	1.94 ± 0.06
Pd-106	0.419 ± 0.014[c]
Te-128	0.372 ± 0.011
La-139	6.62 ± 0.20
Ce-140	6.48 ± 0.19
Pr-141	5.97 ± 0.18

[a] An average value from six U–Mo plates: TUB041, TUB025, TUB021, TUB022, TUB020 and TUB048

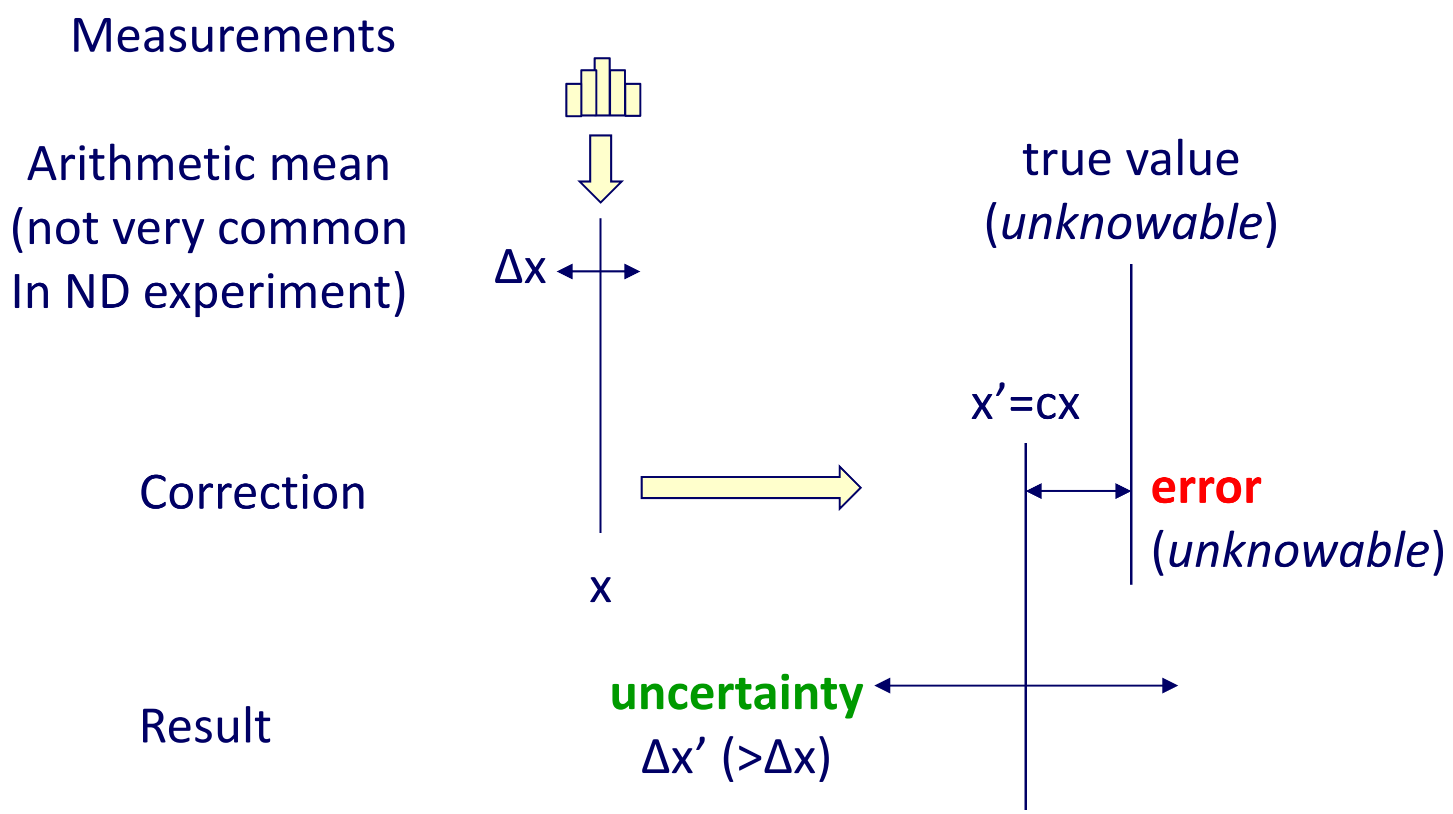
[b] All standard deviations are presented as two sigma values

“sigma” is a synonym of standard deviation.

Probably it be should read as “All uncertainties are presented as two sigma values”.

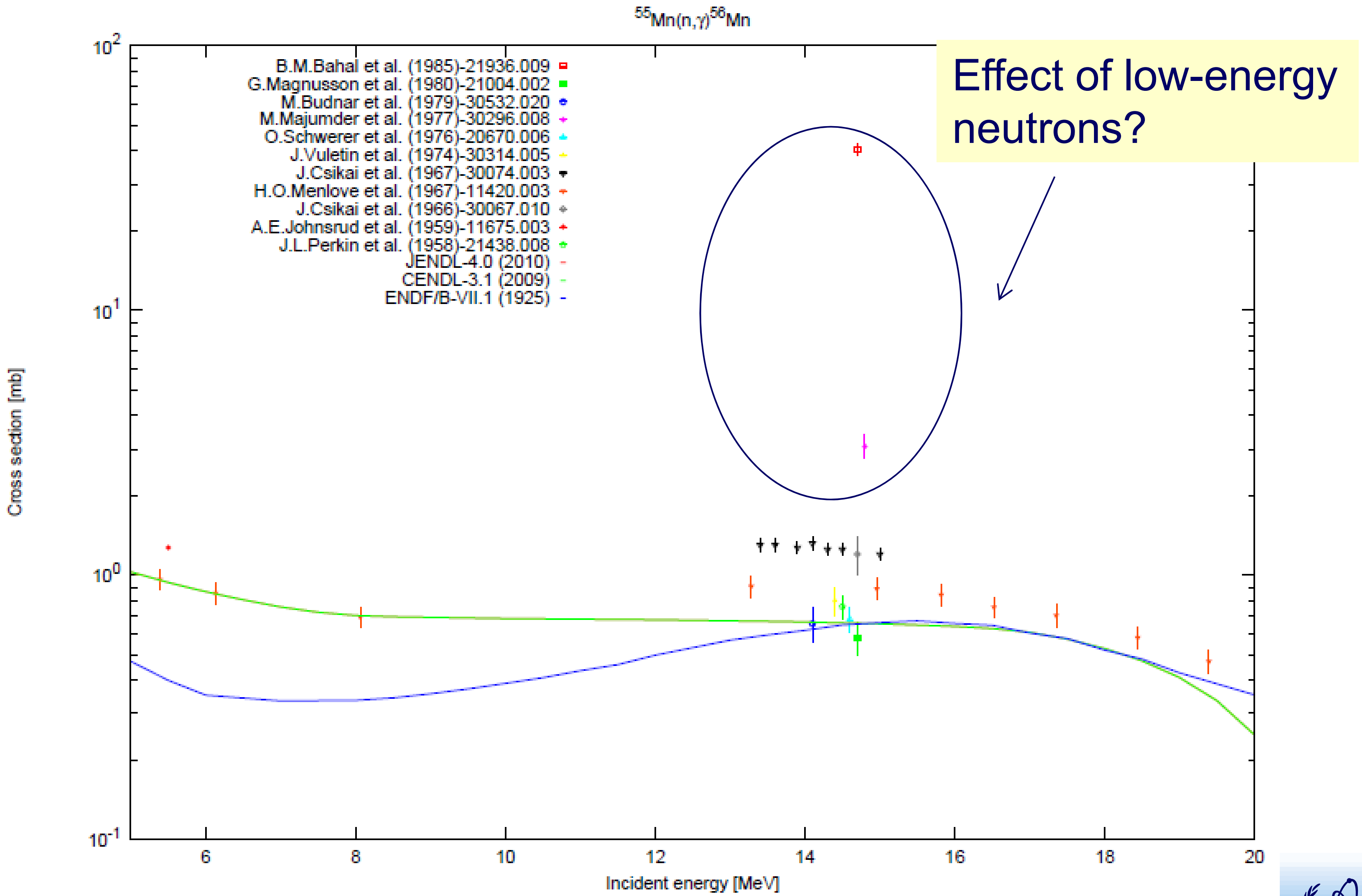
Uncertainty and Error

c.f. Fig.D.2 of GUM2008

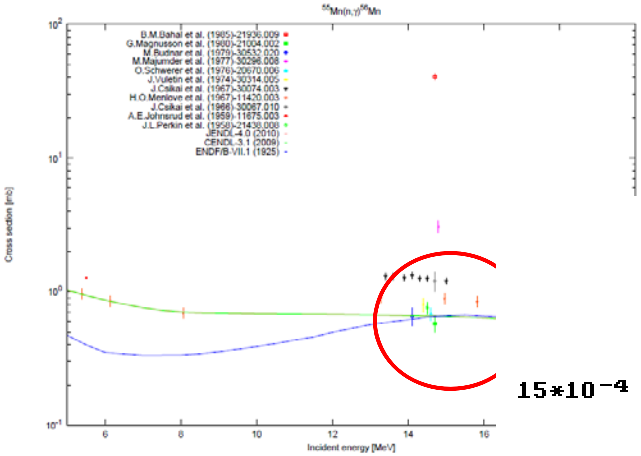


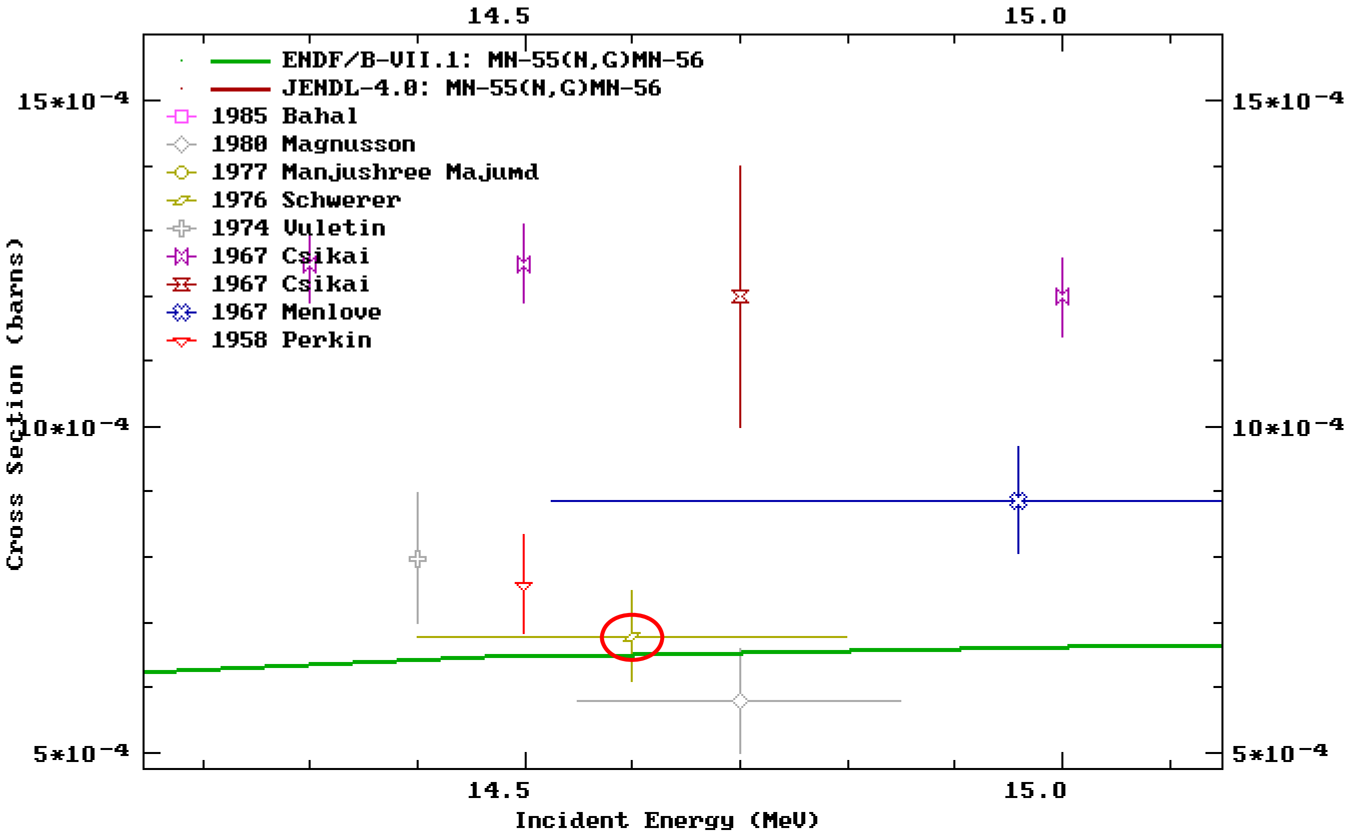
The true value is within the uncertainty.
(successful estimation)

Example of Corrections– $^{55}Mn(n,\gamma)^{56}Mn$



Closer Look Around 14 MeV





Various Corrections for Lower Energy Neutrons

TABLE 1

Essential experimental data and principal features of the decay schemes used for the evaluation of the activation cross sections

Product nucleus	Sample: material [a])	Sample: purity [b]) (%)	Sample: weight (mg)	Decay scheme: γ-transition used for evaluation: energy (keV)	Decay scheme: γ-transition used for evaluation: int. per 100 decays	Decay scheme: half-life	Decay scheme: ref.	Corrections [c]): lower energy neutrons (% of measured activity): thermal, epithermal	Corrections [c]): lower energy neutrons (% of measured activity): prod. in target	Corrections [c]): lower energy neutrons (% of measured activity): prod. in sample	Corrections [c]): γ-absorption in sample (%)
^{38}Cl	$BaCl_2$ p	99.998	658.8	2167	42	37.3 min	[47])		1.8		
^{42}K	KI p	sp	594.5	1524	18.3	12.4 h	[47])		not calculated		
^{51}Ti	Ti, sheet	99.97	115.64	320	95	5.76 min	[48])		1.7		
^{52}V	V_2O_5 p	99	418.25	1134	100	3.75 min	[48])		2.4	1.5	
^{56}Mn	Mn_3O_4 p	99.995	472.25	846.5	98.8	2.582 h	[49])	0.5	5	0.5	1
^{72}Ga	Ga_2O_3 p	99.999	506.65	834.7	95	14.1 h	[50])		13.5	1	1
^{88}Rb	RbCl p	sp	509.8	1836	23.2	17.8 min	[51])		not calculated		
^{90}Y	Y_2O_3 p	99.9999	364.43	480	99.6	3.19 h	[15])		1.5		1
^{128}I	KI p	sp	594.5	442.9	15.8	25 min	[52, 53])	1	36	8.3	2
^{131g}Te	TeO_2 p	99.999	272.35	149.7	80	25 min	[54–56])		6.9	2.0	4
^{131m}Te	TeO_2 p	99.999	272.35	334.5	14.4	1.25 d	[54–56])		not calculated		1
^{139}Ba	$BaCl_2$ p	99.998	658.8	166	22.1	85 min	[57, 51, 53])		2.8		5.7
^{140}La	La_2O_3 p	99.999	440.5	1596	95.3	40.2 h	[51, 58])		7.3	0.4	
^{143}Ce	CeO_2 p	99.9	495.5	293.3	49.5	33.7 h	[59])				2.3
^{187}W	WO_3 p	99.9	591	685.7	28.9	23.8 h	[60])		16.5	5.2	1
^{199}Pt	Pt, sheet	99.97	2161.2	542.7	15.5	31 min	[61, 53])		27	30	8.2
^{198}Au	Au, sheet	99.99	198.75	412	99.82	2.698 d	[62])	3.35	31.9	3.8	2

[a]) The abbreviation p is used for powder.
[b]) The abbreviation sp is used for suprapure, according to definition by Merck Laboratories, Germany, who supplied target materials in these cases. In all other cases target materials were supplied by Koch-Light Laboratories Ltd., England, or by Goodfellow Metals Ltd., England.
[c]) Where no value appears the correction was negligible.

O. Schwerer et al., Nucl. Phys. **A264**(1976)105 (EXFOR 20670)

Note: Correction procedures improve the best estimate, but also introduce a new source of uncertainty.

Poisson Distribution

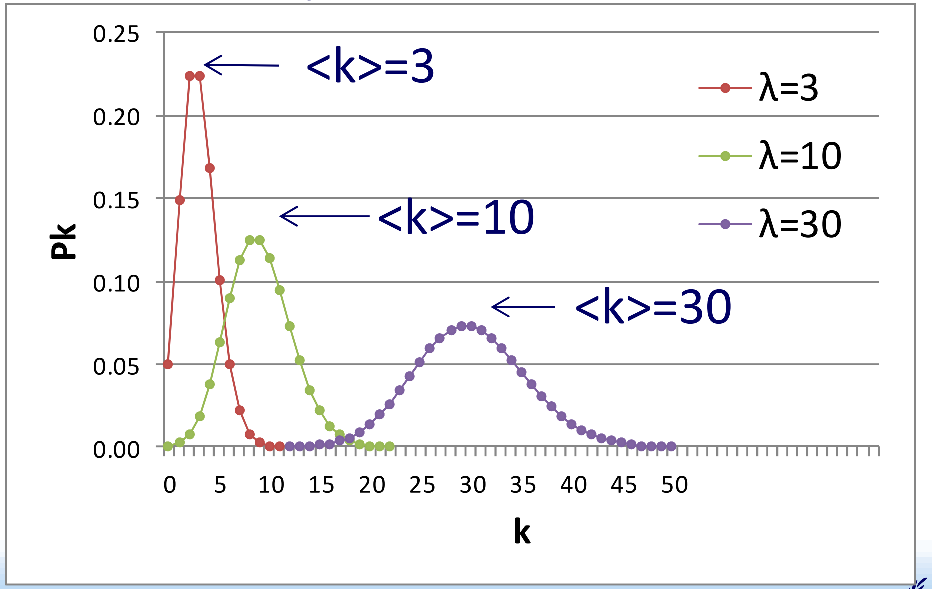
If the event occur

- with a known mean rate (= λ events in a given time span ΔT);
- independently of the time since the last event;
- one time at maximum within an appropriate Δt (<< ΔT);

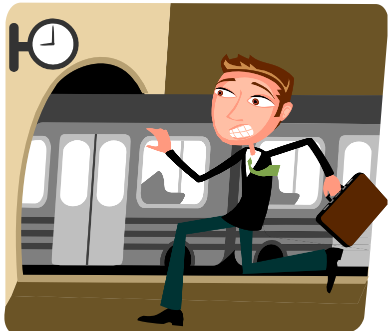
the probability distribution is described by **Poisson distribution**:

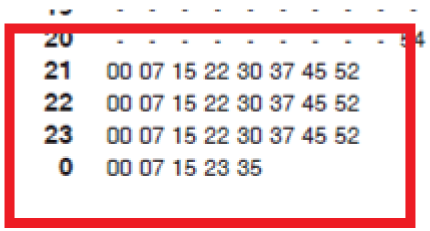
$P_k = \lambda^k \exp(-\lambda)/k!$

Siméon Denis Poisson
(1781 – 1840)



Occurs Independently of the Time Since the Last Event?





20	- - - - - - - - - 54
21	00 07 15 22 30 37 45 52
22	00 07 15 22 30 37 45 52
23	00 07 15 22 30 37 45 52
0	00 07 15 23 35

Occurs dependently of the time since the last subway

You can be more optimistic (?) if the subway is scheduled by Poisson Statistics

Basic Properties of Poisson Distribution

You may prove that Poisson distribution $P_k = \lambda^k \exp(-\lambda)/k!$ satisfies

- normalization $\Sigma_{k=0,\infty} P_k = 1$
- mean $\langle k \rangle = \Sigma_{k=0,\infty} k \cdot P_k = \lambda$
- variance $v = \Sigma_{k=0,\infty} k^2 \cdot P_k - \langle k \rangle^2 = \lambda$
- standard deviation $\Delta k = (v)^{1/2} = \lambda^{1/2}$

Hint: $\Sigma_{n=0,\infty} \lambda^n / n! = e^\lambda$

mean = variance!

Poisson Distribution: Counting Experiment

Suppose we repeated counting experiment n times, and obtain count N_i (i=1,,n). If the phenomenon follows the Poisson distribution, we obtain

- mean $\langle N\rangle = (\Sigma_{i=1,n} N_i)/n$
- variance $v = (\Sigma_{i=1,n} N_i^2) / n - \langle N\rangle^2 \sim \langle N\rangle$
- standard deviation $\Delta N = v^{1/2} \sim \langle N\rangle^{1/2}$

In nuclear reaction measurements, count N from a single counting measurement is often assumed to be <N> (i.e., $N \sim \langle N\rangle$, $\Delta N \sim \langle N\rangle^{1/2}$).

Poisson distribution: mean $\langle k\rangle = \lambda$, variance $v = \lambda$, standard deviation $s = \lambda^{1/2}$.

Exercise: Irradiation Time

- We want to determine $Y=N/\varepsilon$ by measuring count N with a detector (efficiency ε, $\Delta\varepsilon/\varepsilon=5.0\%$). We expect 600 counts/min.
- Fill the table below by assuming Poisson statistics (i.e., $\Delta N=N^{1/2}$) and error propagation $(\Delta Y/Y)^2=(\Delta N/N)^2+(\Delta\varepsilon/\varepsilon)^2$.
- How long should we irradiate the sample? (c.f. DS12 Table VI).

N	ΔN/N (%)	Δε/ε (%)	ΔY/Y (%)	time (min)
100		5.0		
500		5.0		
1,000		5.0		
10,000		5.0		
20,000		5.0		
50,000		5.0		
100,000		5.0		
200,000		5.0		

Exercise: Irradiation Time

N	ΔN/N (%)	Δε/ε (%)	ΔY/Y (%)	time (min)
100	10.0	5.0	11.2	0.2
500	4.5	5.0	6.7	0.8
1,000	3.2	5.0	5.9	1.7
10,000	1.4	5.0	5.2	8.3
20,000	1.0	5.0	5.1	16.7
50,000	0.4	5.0	5.0	83.3
100,000	0.3	5.0	5.0	166.7
200,000	0.2	5.0	5.0	333.3

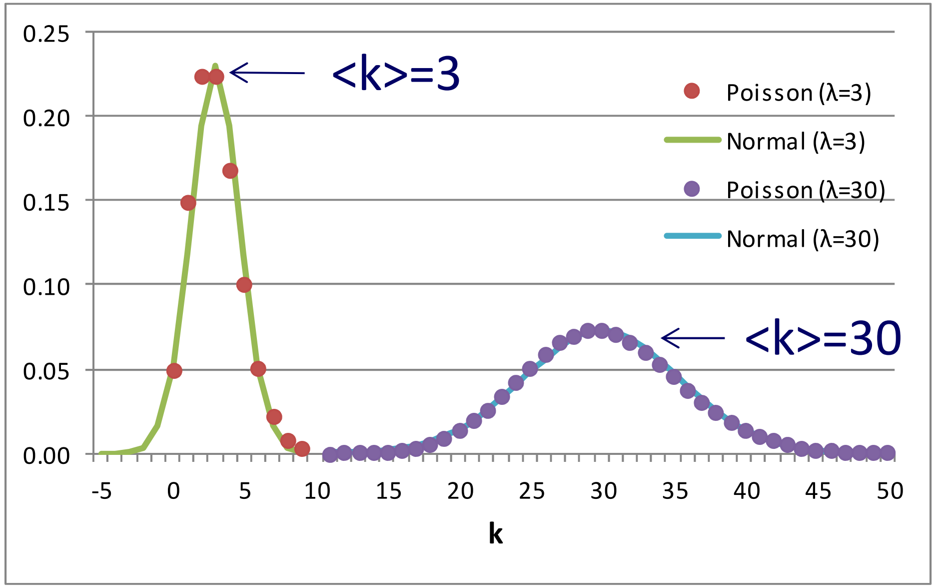
Counting more than ~100 min does not improve the uncertainty!

Normal (Gauss) Distribution

For an enough large mean number, the Poisson distribution is well approximated by the **normal (Gauss) distribution**

$$P_k = \lambda^k \exp(-\lambda)/k! \rightarrow P(k) = \exp[-(k-\lambda)^2/(2\lambda)] / (2\pi\lambda)^{1/2}$$

Johann Carl Friedrich Gauß (1777 – 1855)



Note that P(k) is a probability <u>density</u> distribution. The probability to find a value k within $[x_{min}, x_{max}]$ is $P_k \sim \int_{xmin}^{xmax} dx\, P(x)$.

Properties of Normal (Gauss) Distribution

You may prove that normal distribution, $P(x) = \exp[-(x-\lambda)^2/(2\lambda)] / (2\pi\lambda)^{1/2}$ satisfies that

- Normalization $\int_{-\infty}^{+\infty} dx\, P(x) = 1$
- mean $x_0 = \langle x \rangle = \int_{-\infty}^{+\infty} dx\, x \cdot P(x) = \lambda$
- variance $v = \int_{-\infty}^{+\infty} dx\, x^2 \cdot P(x) - \langle x \rangle^2 = \lambda$
- standard deviation $\Delta x = (v)^{1/2} = \lambda^{1/2}$

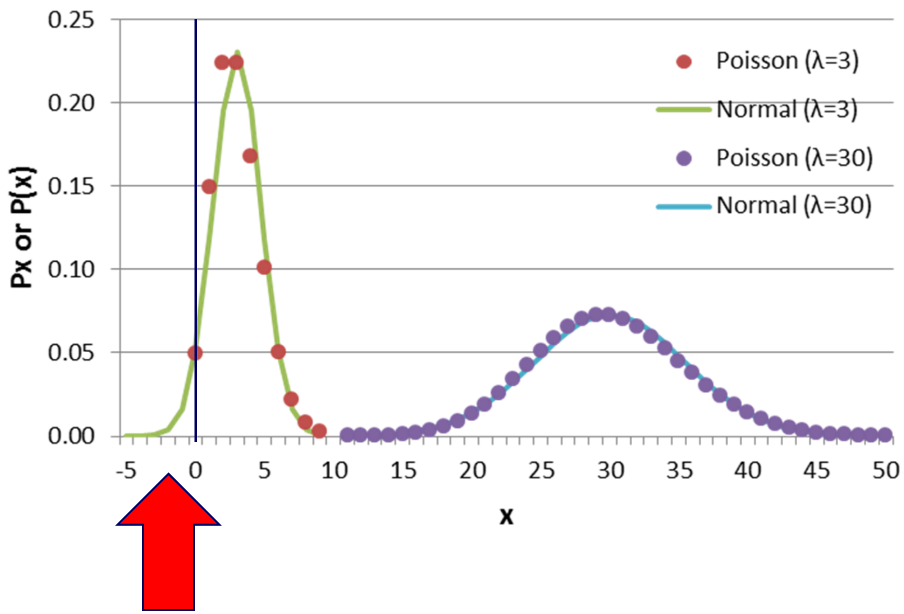
Hint: $\int_{-\infty}^{+\infty} dx \exp(-ax^2) = (\pi/a)^{1/2}$,

$\int_{-\infty}^{+\infty} dx\, x \cdot \exp(-ax^2) = 0$,

$\int_{-\infty}^{+\infty} dx\, x^2 \cdot \exp(-ax^2) = (\pi^{1/2})/(2a^{3/2})$

Remarks on Normal Distribution

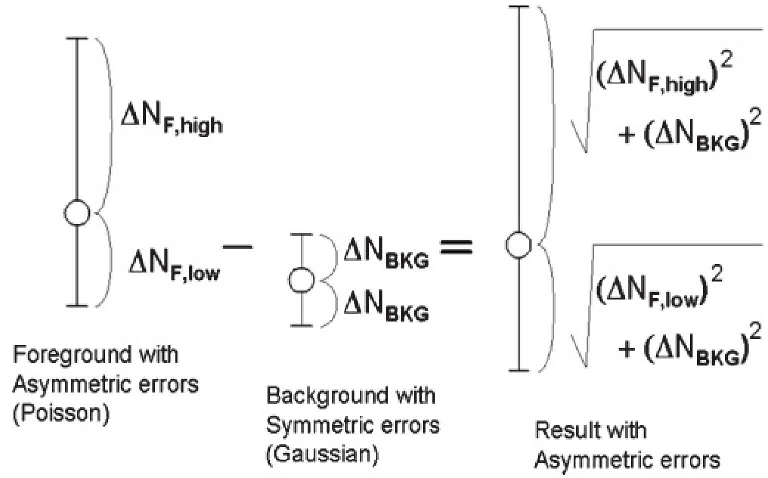
Poisson distribution P_x is defined for a non-negative random variables x. However normal distribution P(x) may give finite probability for <u>negative</u> x.

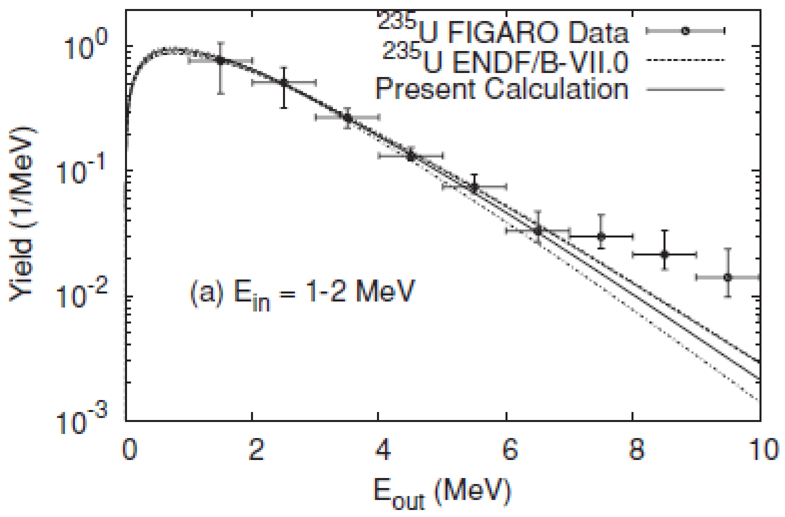


The normal distribution is a good approximation of the Poisson distribution when we have enough counting number x.

An Example of Poisson Distribution in EXFOR

In our experiments, the foreground statistics are poor, and there are some zero counts in the energy-binned foreground spectra so that it may be inappropriate to assume Gaussian distributions for the data. The error bars of foreground events are estimated with ROOFIT [27] using the Poisson distribution, which is not symmetric. Here we describe the upper side of the error bar, $\Delta N_{\mathrm{F,\,high}}$, and the lower side, $\Delta N_{\mathrm{F,\,low}}$.



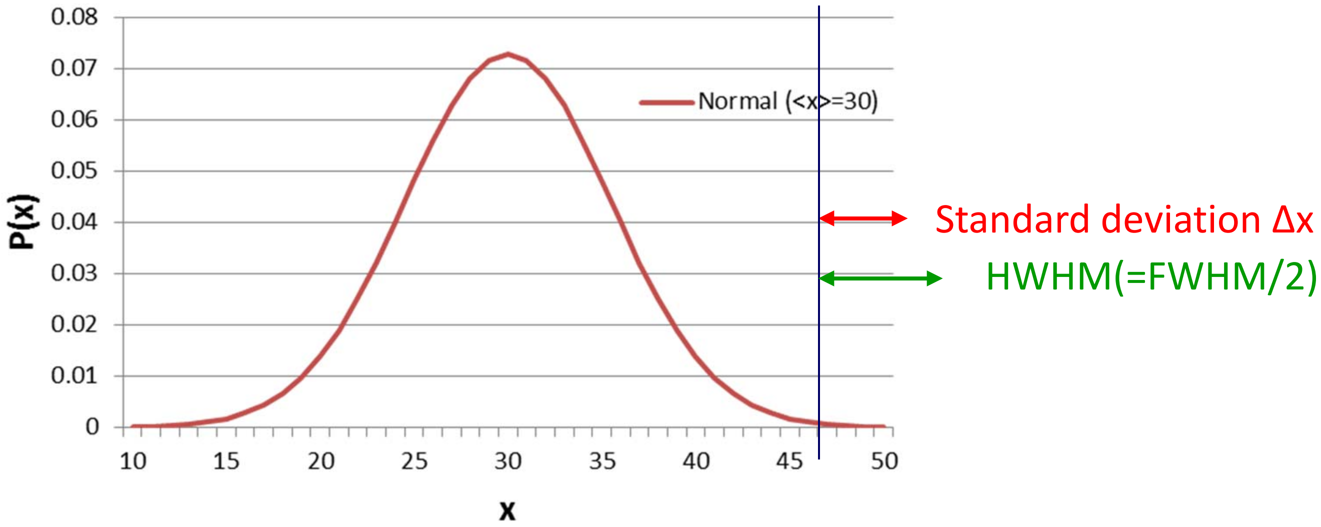



```
SUBENT        14290002   20111006
BIB                  3          7
REACTION   (92-U-235(N,F),PR,NU/DE,,AV/REL)
            Spectra are normalized to unity by integrating the
             spectra in the 2.0-6.5 MeV emission energy range
ERR-ANALYS (ERR-S) Statistical uncertainty, considering
                   - foreground counting statistics  (Poisson)
                   - background counting statistics (Gaussian)
STATUS     (TABLE) Plotted in Fig.7 of Phys.Rev.C83(2011)034604
ENDBIB               7
NOCOMMON             0          0
DATA                 6         63
EN-MIN      EN-MAX      E          DATA       -ERR-S     +ERR-S
MEV         MEV         MEV        ARB-UNITS  ARB-UNITS  ARB-UNITS
 1.          2.            1.5      7.6634E-01 3.4578E-01 2.9948E-01
 1.          2.            2.5      5.1220E-01 1.9031E-01 1.6367E-01
 1.          2.            3.5      2.7136E-01 4.7978E-02 4.5734E-02
```

S. Noda et al., Phys. Rev. C 83(2011)034604 (EXFOR 14290)

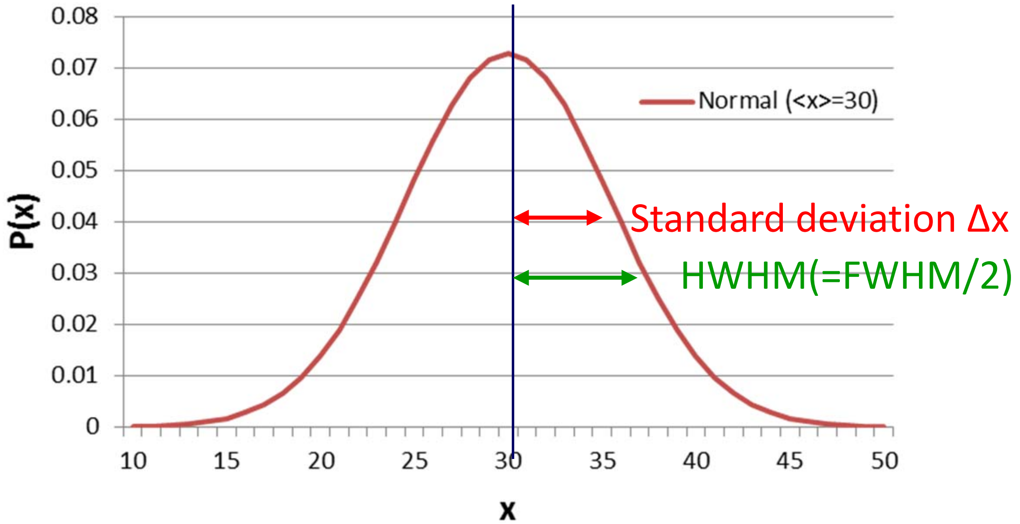
Normal Distribution:Standard Deviation (SD) and HWHM

Each result falls within <x>± SD in 68% probability (confidential level) in the normal distribution (SD: standard deviation s).



By using the definition of the normal distribution, you can easily prove that the full width at half maximum (HWHM) = (2 ln 2) Δx ~ 1.2 Δx.

Various Error Measures in LEXFOR



4. **error measure**, such as:

- standard deviation ≠ half-width at half-maximum of Gaussian error distribution function
  = 2/3 probability that the true value is within error bars
- confidence limits: when errors are given as confidence limits various definitions exist, for example, 95% probability, which corresponds to approximately two standard deviations.
- errors supposed not to exceed: approx. 100% probability value is within error bars.

Multiple Random Variables

As an extension of the probability distribution for a single random variable P_k or P(x), we can consider the distribution for two more random variables $P_{k,l,m}$,… or P(x,y,z,…).

Example: Probability to see k on the first dice and see l on the second dice

k	1	1	1	1	1	1	2	…
L	1	2	3	4	5	6	1	…
$P_{k,l}$	1/36	1/36	1/36	1/36	1/36	1/36	1/36	…

Mean, Variance and Standard Deviation (Discrete Multiple Variable)

Definition of mean, variance and standard deviation for a discrete random variable k following the probability distribution $P_{k,l,}$:

- **Mean**

  $<k> = \Sigma_{k=1,n} k \cdot P_{k,l,\ldots}$

- **Variance**

  $v_{kk} = <(k-<k>)^2> = <dk \cdot dk>= <k^2>-<k>^2$

- **Standard deviation**

  $\Delta k=(v)^{1/2}$

- **Covariance** (not defined in single random variable distribution)

  $v_{kl} = <(k-<k>)(l-<l>)>=<k \cdot l>-<k><l>$

Mean, Variance and Standard Deviation (Continuous Variable)

For continuous multiple random variable x, y,… similarly

- **Mean**

  $<x> = \int dx\, x \cdot P(x,y,\ldots)$

- **Variance**

  $v_x = <(x-<x>)^2> = <x^2>-<x>^2$

- **Standard deviation**

  $\Delta x=(v_x)^{1/2}$

- **Covariance**

  $v_{xy}= <(x-<x>)(y-<y>)>=<x \cdot y>-<x><y>$

Maxwell Distribution

Probability to find an atom with velocity (v_x,v_y,v_z) in equilibrium:

$$P(v_x,v_y,v_z)=[m/(2\pi kT)]^{3/2}\exp[-m(v_x^2+v_y^2+y_z^2)/(2kT)]$$

This can be decomposed to the distribution for x, y and z:

$P(v_x,v_y,v_z)=p(v_x)p(v_y)p(v_z)$ with normal distribution for single variable

$p(v_x)=[m/(2\pi kT)]^{1/2}\exp[-mv_x^2/(2kT)]$ etc. $\int dv_i\, p(v_i) =1$ (i=x,y,z)

- mean $\langle v_x\rangle = \int dv_x dv_y dv_z\, v_x\cdot P(v_x,v_y,v_z)$

  $=\int dv_x\, v_x\cdot p(v_x) \int dv_y p(v_y)\cdot\int dv_z p(v_z) =0$

- variance $v_{xx}=\langle v_x^2\rangle-\langle v_x\rangle^2$

  $=\int dv_x dv_y dv_z\, v_x^2\cdot P(v_x,v_y,v_z)-\langle v_x\rangle^2$

  $=\int dv_x\, v_x^2\cdot p(v_x) \int dv_y p(v_y)\cdot\int dv_z p(v_z)=kT/m$

Covariance

For probability density distribution $P(x,y,z,\ldots)$,

$$\mathbf{V_{xy} = \langle (x-\langle x\rangle)(y-\langle y\rangle)\rangle = \langle xy - x\langle y\rangle - \langle x\rangle y + \langle x\rangle\langle y\rangle\rangle = \langle xy\rangle - \langle x\rangle\langle y\rangle}$$

is defined as the **covariance**. V_{xx} gives the variance of x.

When two random variables x and y are independent, e.g.,

$P(x,y) = P_x(x)P_y(y) \rightarrow V_{xy} = 0$ (V becomes diagonal).

Example: Maxwell distribution

$$P(v_x,v_y,v_z) = [m/(2\pi kT)]^{3/2} \exp[-m(v_x^2+v_y^2+y_z^2)/(2kT)]$$
$$= \Pi_{i=x,y,z}\, [m/(2\pi kT)]^{1/2} \exp[-mv_i^2/(2kT)]$$

(Motions for x-, y- and z- directions are <u>independent</u> each other.)

Vector Notation of Random Variable

x: random variable

x_i: i-th sample of x (i=1,n)

n samples of single random variable:

$\mathbf{x}=\{x_i\}$ (i=1,n)

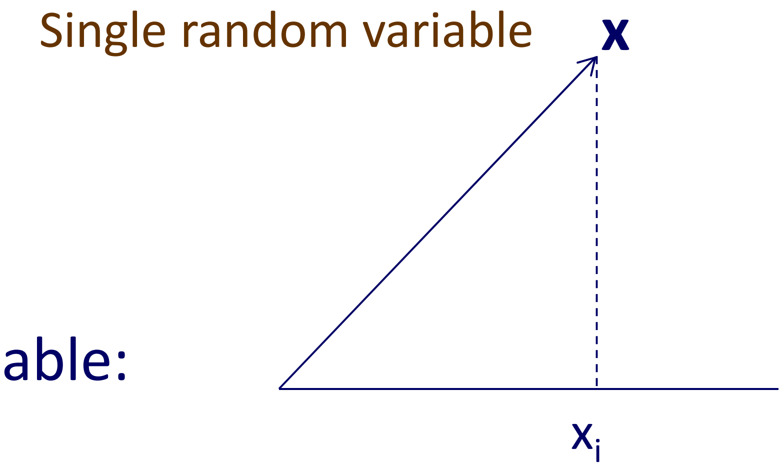
in the vector notation.

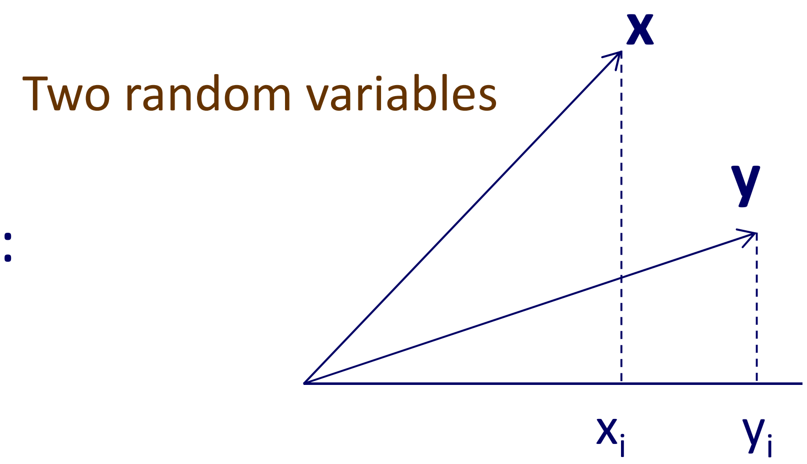
n samples of multi random variables:

$\mathbf{x}=\{x_i\}$ (i=1,n)

$\mathbf{y}=\{x_i\}$ (i=1,n)

…

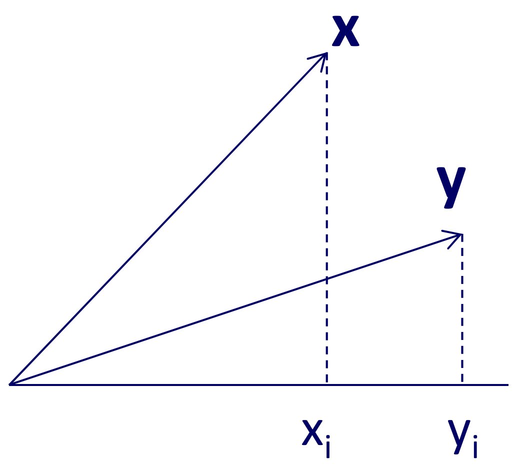




Vector Notation of Random Variables

x, y: random variable

x_i and y_i: i-th sample of x and y (i=1,n)



$<x^2> = \sum_{i=1,n} x_i^2/n = |\mathbf{x}|^2/n$

variance

$V_{xx} = <(x-<x>)^2>$

$= [\sum_{i=1,n} (x_i - <x>)(x_i - <x>)]/n$

$= \mathbf{X}\cdot\mathbf{X} = |\mathbf{X}|^2$

covariance

$V_{xy} = <(x-<x>)(y-<y>)>$

$= [\sum_{i=1,n} (x_i - <x>)(y_i - <y>)]/n$

$= \mathbf{X}\cdot\mathbf{Y}$

$\mathbf{a}\cdot\mathbf{b} = \sum_{i=1,n} a_i b_i$ (inner product)

$\mathbf{a}\cdot\mathbf{a} = |\mathbf{a}|^2$ (length of vector)

New variable:

$\mathbf{X} = \{X_i\}$ with $X_i = (x_i - <x>) n^{1/2}$

Correlation Coefficient

For variance V_{xx}, V_{yy} and covariance V_{xy} ,

$$\mathbf{c_{xy} = V_{xy} / (V_{xx} \cdot V_{yy})^{1/2}}$$

is defined as the **correlation coefficient**.

By using the vector notation, $V_{xx} = |\mathbf{X}|^2$, $V_{yy} = |\mathbf{Y}|^2$, $V_{xy} = \mathbf{X} \cdot \mathbf{Y}$

$$c_{xy} = V_{xy} / (V_{xx} \cdot V_{yy})^{1/2}$$

$$= (\mathbf{X} \cdot \mathbf{Y}) / (|\mathbf{X}| \cdot |\mathbf{Y}|)$$

$$= \cos \theta_{xy}.$$

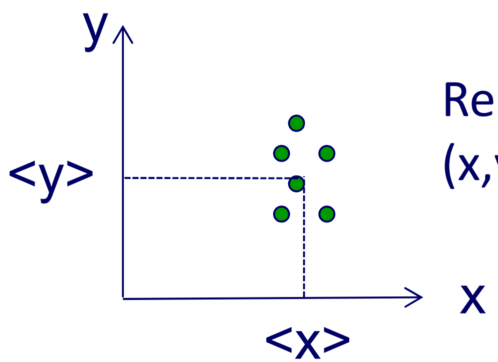
→ $\mathbf{-1 \le c_{xy} \le 1}$

Random variable X: $X=\{X_i\}$, Variance $V_{xx} = \mathbf{X} \cdot \mathbf{X}$, Covariance $V_{xy} = \mathbf{X} \cdot \mathbf{Y}$

Uncorrelated and Fully Correlated Parameters

Six measurements of two uncorrelated parameters (x,y)

$P(x,y) = P_x(x)\, P_y(y)$ (x and y are <u>independent</u>) → $c_{xy}=0$



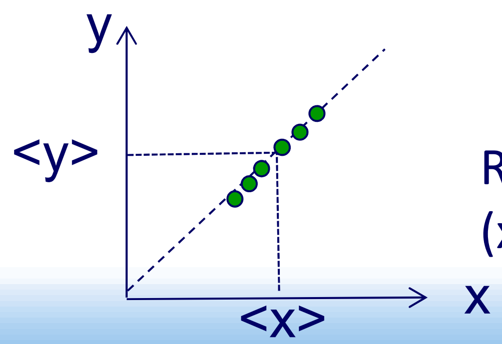
Results of i-th measurement
$(x,y)=(x_i,y_i)$

Six measurements of two fully correlated parameters (x,y)

$P(x,y) = P_x(x)\delta(x-y)$ (x and y are <u>same</u>) → $c_{xy}=1$

$\delta(x-y)$: Dirac's delta function

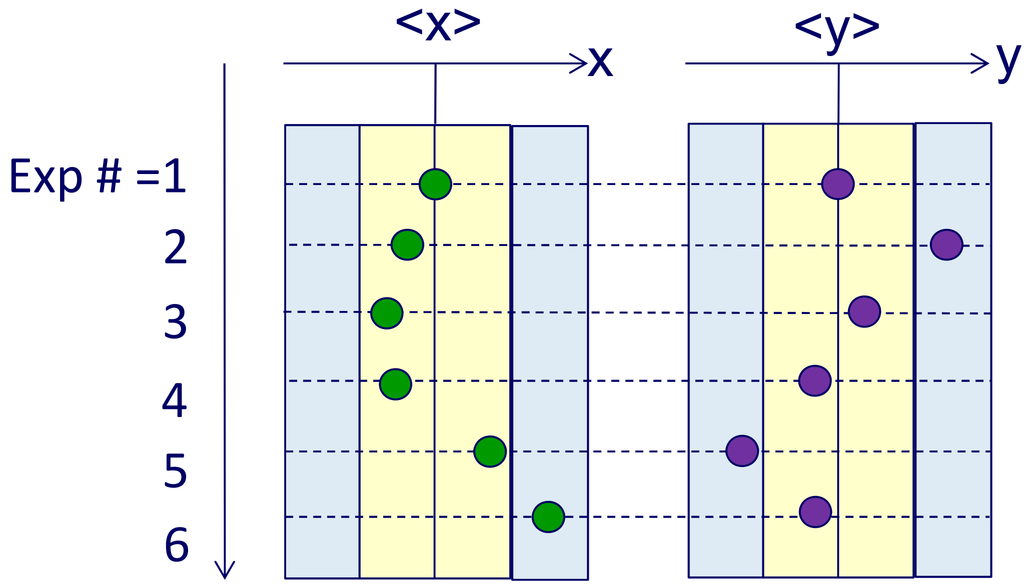
- $\delta(x-y)=0$ if $x \neq y$.
- $\int dx\, \delta(x-y)=1$
- $\int dx\, \delta(x-y) f(y)=f(x)$



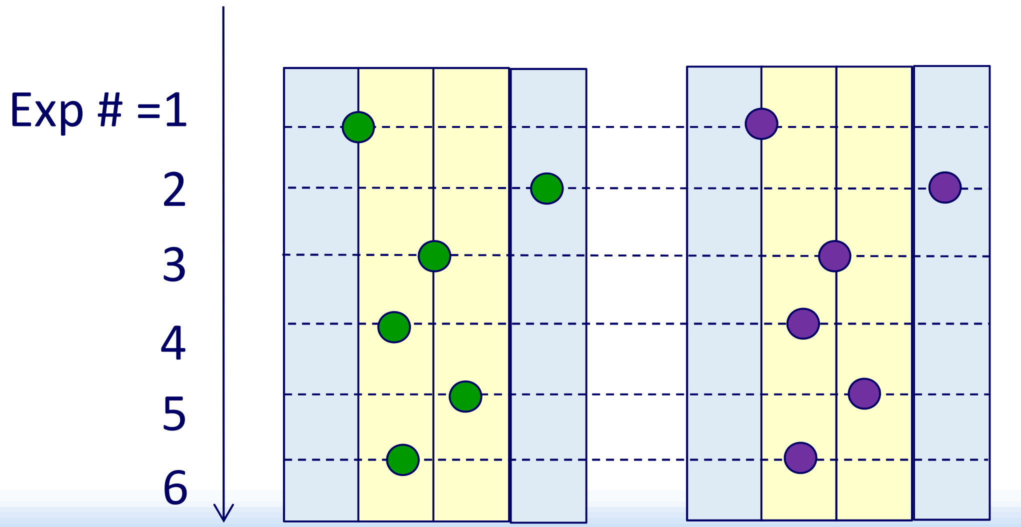
Results of i-th measurement
$(x,y)=(x_i,y_i)$

Correlation between Two Observables (x, y)

Uncorrelated (c_{xy}=0)

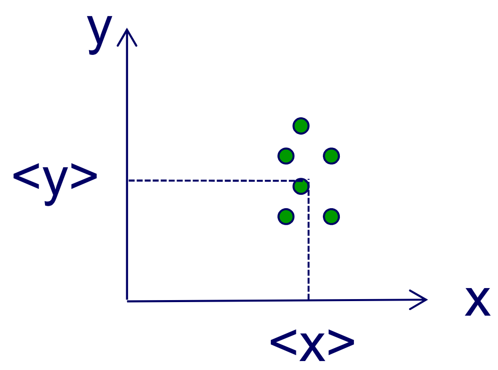


Fully correlated (c_{xy}=1)

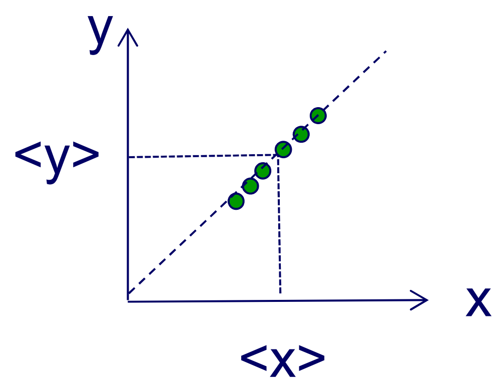


Correlation is not Property of Nature

Two random variables behave independent if they are measured independently.



Correlation between them appears due to by a procedure is introduced by experimental procedure, e.g.,



- x=y assumed by measuring only x (or y);

-x and y derived from correlated two observables p and q (e.g., interpolation from fitting)

Summary

- Direct observables are random variables.
- Standard deviation (square root of variance) is often adopted as the “uncertainty”.
- Poisson distribution: Mean=$\langle N \rangle$, Uncertainty=$\langle N \rangle^{1/2}$
- $\langle N \rangle = N$, $\Delta N = \langle N \rangle^{1/2}$ are often done from a single measurement.
- If N is enough large, Poisson distribution →normal distribution.
- For multiple random variables, covariance is defined.
- Correlation coefficient $c_{xy}=V_{xy}/(V_{xx}V_{yy})^{1/2}$.
- Fully correlated $c_{xy}=1$, uncorrelated $c_{xy}=0$.
- Correlation is introduced by experimentalists!