International Atomic Energy Agency

Multiple Random Variables and Their Correlation

Naohiko Otsuka

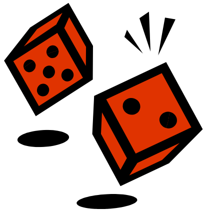
IAEA Nuclear Data Section

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 $P_{k,l,m,\ldots}$ 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**

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

- **Variance**

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

- **Standard deviation**

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

- **Covariance**

  $v_{xy} = \langle (x-\langle x \rangle)(y-\langle y \rangle) \rangle = \langle x \cdot y \rangle - \langle x \rangle \langle y \rangle$

Correlation Coefficient

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

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

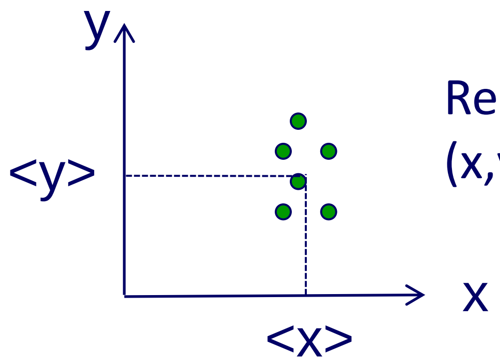
is defined as the **correlation coefficient**.

- $c_{xy}=0$ if x and y are independent.
- $c_{xy}=\pm 1$ if $x=\pm y$.
- In general $-1 \le c_{xy} \le 1$.

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>) → $c_{xy}=0$



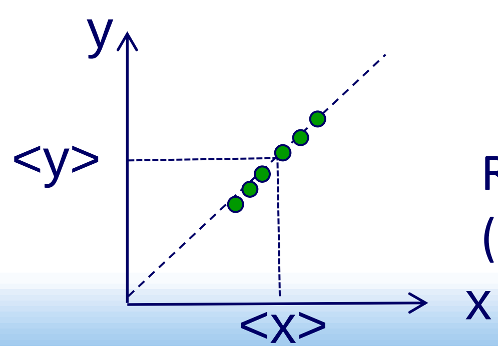
Results of i-th measurement
$(x,y)=(x_i,y_i)$

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>) → $c_{xy}=1$

δ(x-y): Dirac's delta function

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

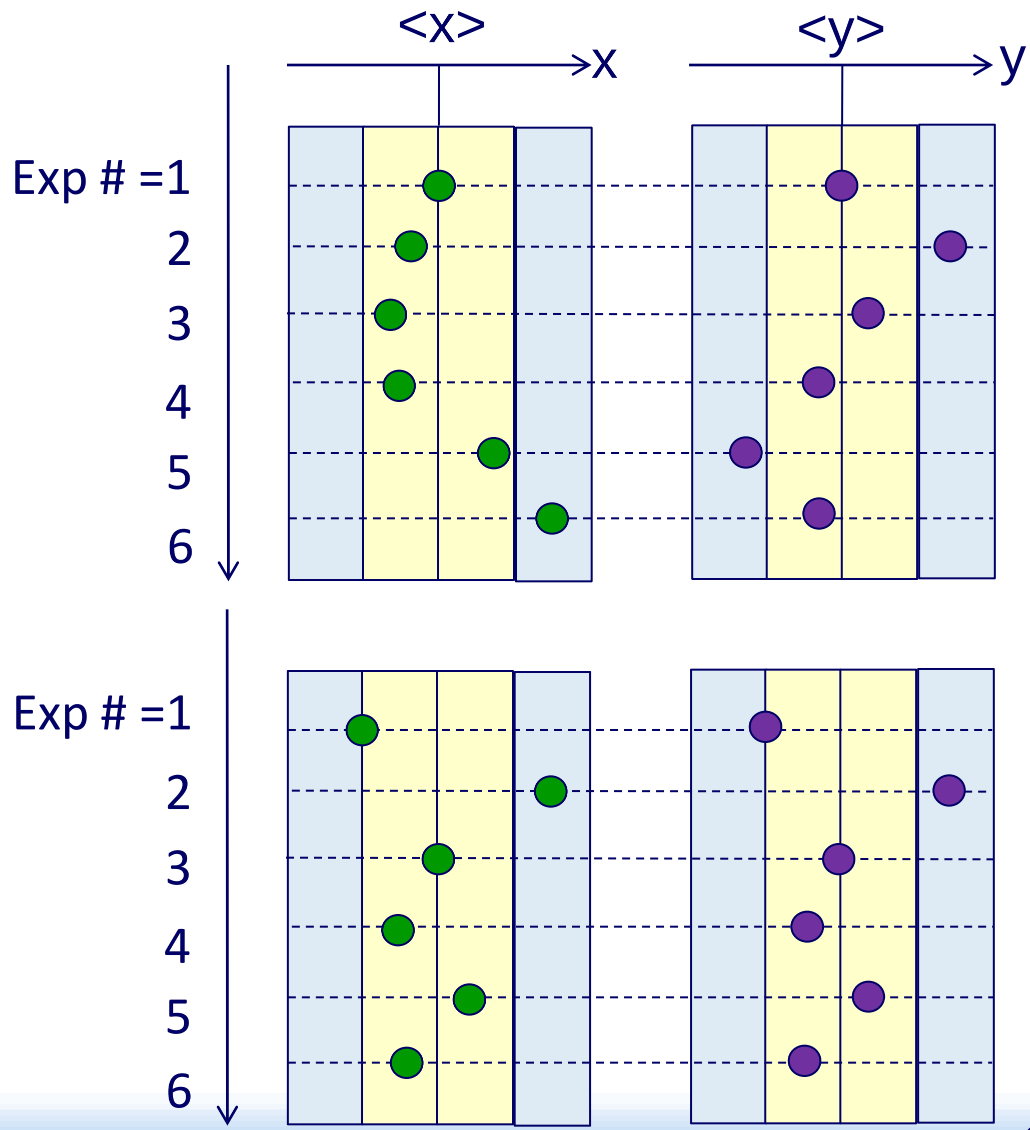


Results of i-th measurement
$(x,y)=(x_i,y_i)$

Correlation between Two Observables (x, y)

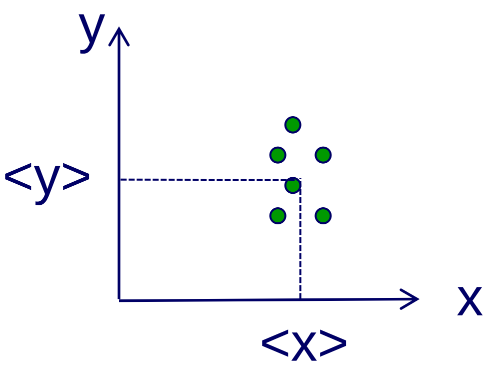
Uncorrelated (c_{xy}=0)

Fully correlated (c_{xy}=1)



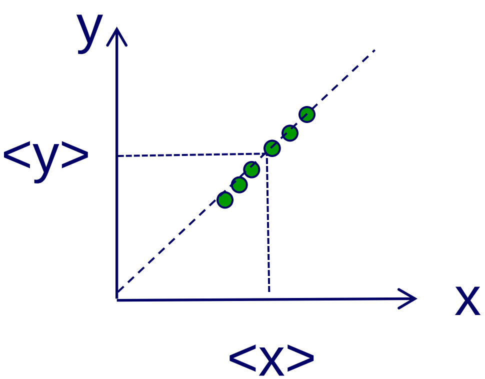
Correlation is *not* Property of Nature

Two random variables behave <u>independent</u> if they are measured *independently*.



Correlation between random variables appears due to a procedure introduced *by experimental procedure*, e.g.,

- x=y assumed by measuring only x (or y);
- x(p,q) and y(p,q) derived from correlated two observables p and q (e.g., interpolation from fitting)



Correlation: Two Reactions and Two Energies

$\sigma = A/(\varepsilon N \phi)$

$\sigma_{xi} = A_{xi}/(\varepsilon_x \cdot N_x \cdot \phi_i)$

$\sigma_{xj} = A_{xj}/(\varepsilon_x \cdot N_x \cdot \phi_j)$

$\sigma_{yi} = A_{yi}/(\varepsilon_y \cdot N_y \cdot \phi_i)$

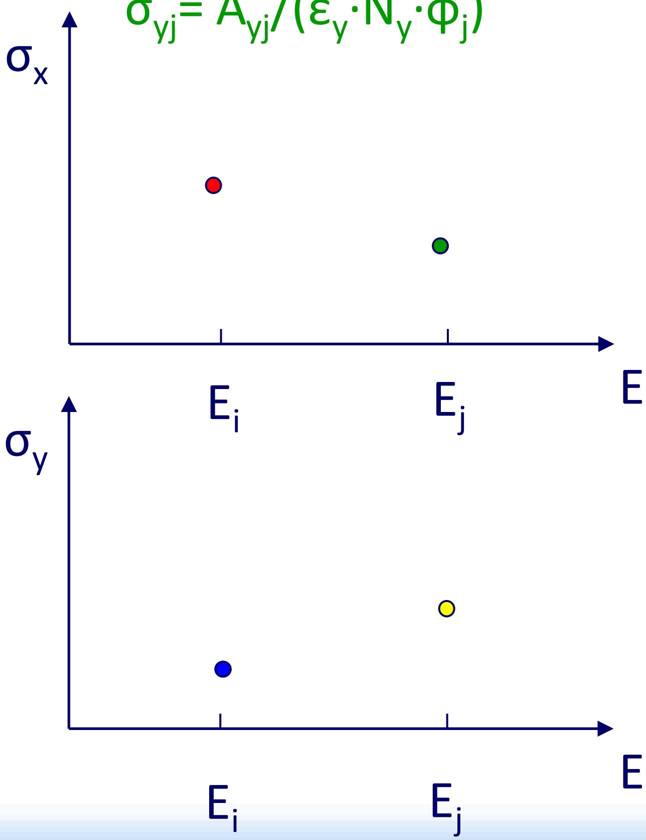
$\sigma_{yj} = A_{yj}/(\varepsilon_y \cdot N_y \cdot \phi_j)$

A: measured counting rate (independent)

ε: detector efficiency (2 values for x and y)

N: number of sample atoms (2 values for x and y)

ϕ: beam flux density (2 values for i and j)



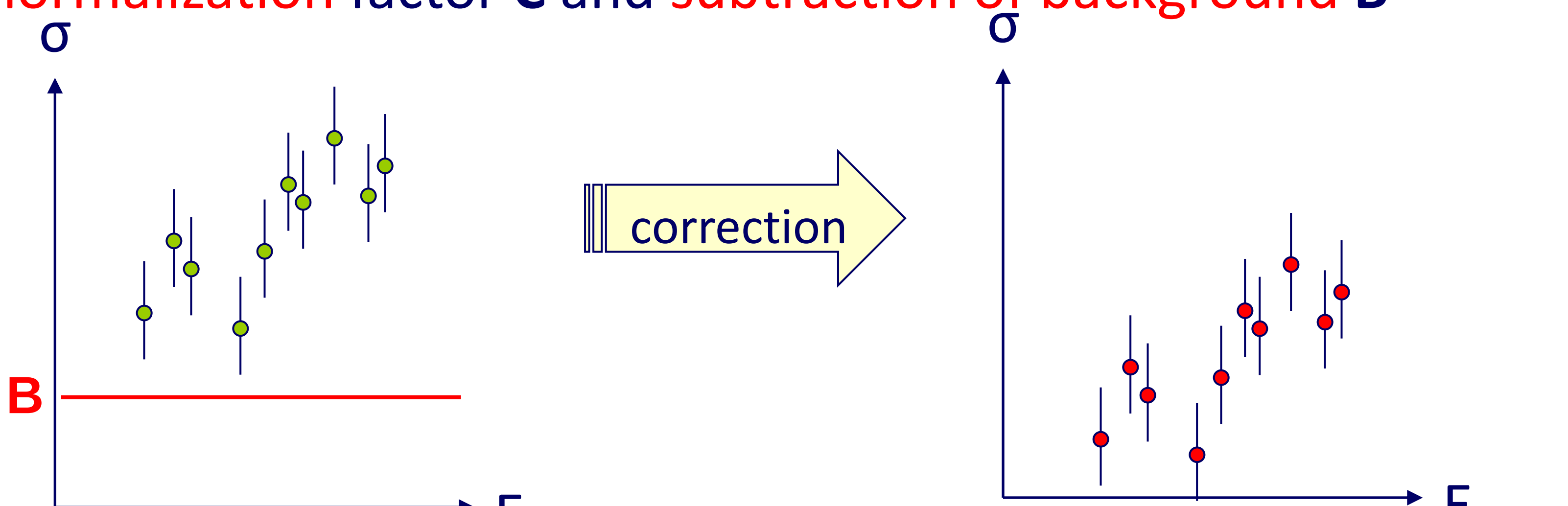
Source of correlation
between two cross sections
due to assumption of equality

	σ_{xi}	σ_{xj}	σ_{yi}	σ_{yj}
σ_{xi}	-	ε_x, N_x	ϕ_i	(ind.)
σ_{xj}		-	(ind.)	ϕ_j
σ_{yi}			-	ε_y, N_y
σ_{yj}				-

ind.: independent

Correlation within Single Experiment

Example: Cross section measurement at various beam energy with renormalization factor **C** and subtraction of background **B**



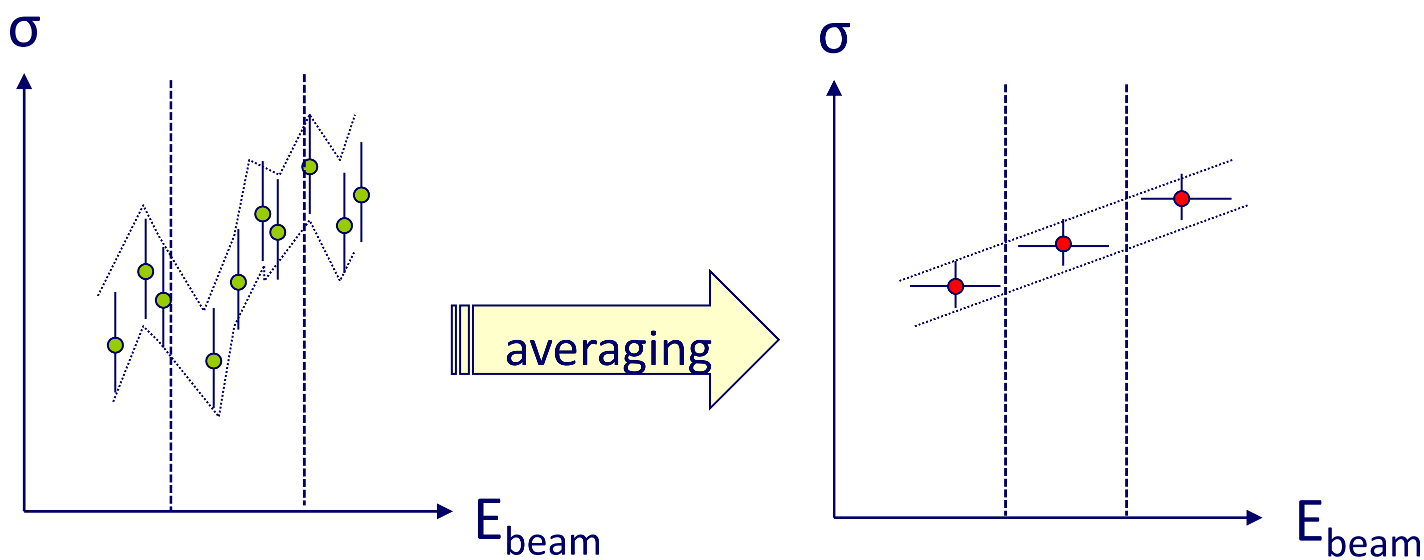
$\sigma_i = N_i / (n_i \rho_i)$

$\sigma_i = (N_i - B) / (n_i \rho_i)$

Typical correlated source within one experiment

- Common sample characterization(n)
- Common normalization factor (C)
- Common background subtraction (B)

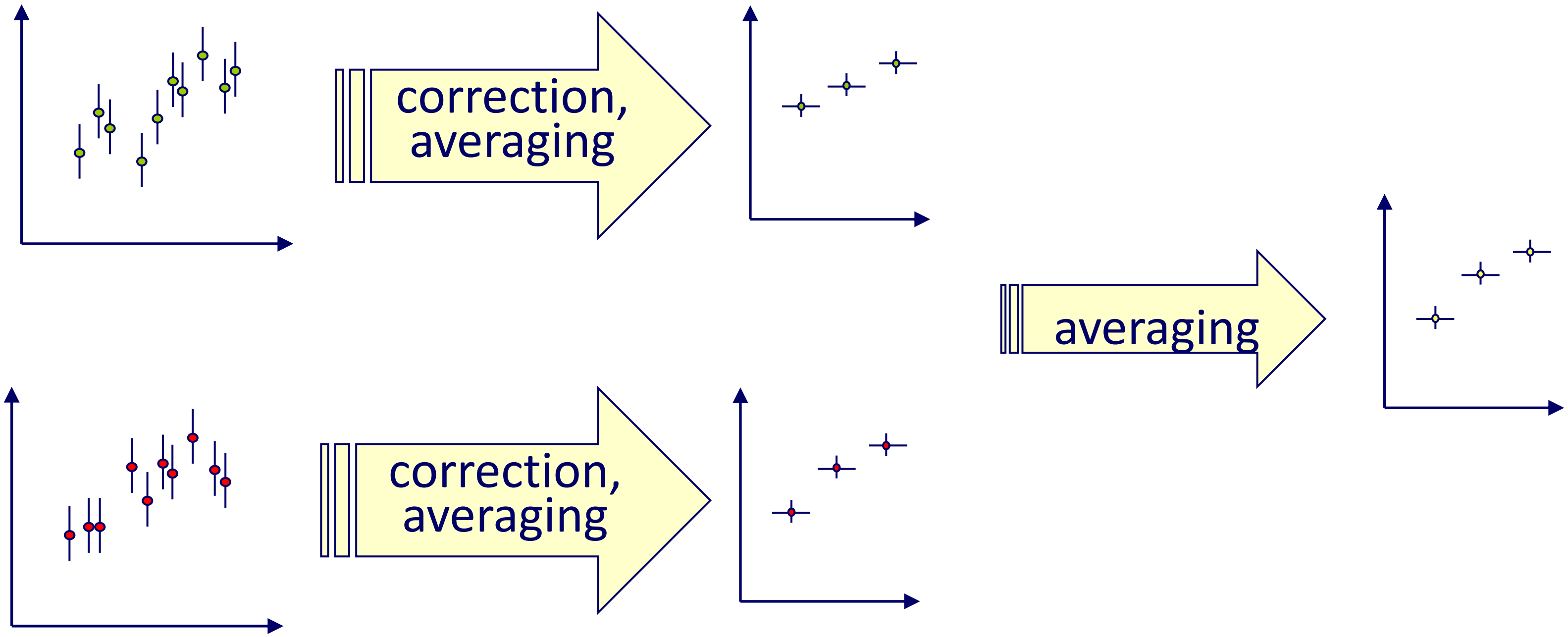
Average within Single Experiment



Statistical fluctuation visible.
Uncorrelated uncertainty dominant.

Statistics (fluctuation) improved.
Correlated uncertainty dominant.

Correlation in Various Steps of Data Reduction



Correlation depends on the data reduction procedure.
(i.e., Correlation is not a physical quantity and not unique.)

Summary

- 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!