Propagating uncertainties

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

Propagating uncertainties

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

"Two and two equal four.

Propagating uncertainties

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

"Two and two equal four.
The trouble is, in that world of shadows and distorting mirrors

Propagating uncertainties

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

"Two and two equal four.
The trouble is, in that world of shadows and distorting mirrors
what may or may not appear to be two,
when multiplied by a factor that may or may not be two,

Propagating uncertainties

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

"Two and two equal four.
The trouble is, in that world of shadows and distorting mirrors
what may or may not appear to be two,
when multiplied by a factor that may or may not be two,
could possibly come out at four but probably will not."

(Frederick Forsyth – *The fist of God*)

Propagation of uncertainties

All we have seen so far in this short review of 'direct probability' is how to 'propagate probability' to logically connected events or variables.

Propagation of uncertainties

All we have seen so far in this short review of 'direct probability' is how to 'propagate probability' to logically connected events or variables.

$\Rightarrow$ Therefore, the famous problem of propagation of uncertainty is straightforward in a probabilistic approach: just use probability theory.

[Note that in the frequentistic approach one does something similar, but in a 'strange' way, because one is not allowed to use probability for values of physical quantities, but only for *estimators*.]

Propagation of uncertainties

All we have seen so far in this short review of 'direct probability' is how to 'propagate probability' to logically connected events or variables.

⇒ Therefore, the famous problem of propagation of uncertainty is straightforward in a probabilistic approach: just use probability theory.

[Note that in the frequentistic approach one does something similar, but in a 'strange' way, because one is not allowed to use probability for values of physical quantities, but only for *estimators*.]

The general problem:

$$f(x_1, x_2, \ldots, x_n) \xrightarrow[Y_j = Y_j(X_1, X_2, \ldots, X_n)]{} f(y_1, y_2, \ldots, y_m)\,.$$

This calculation can be quite challenging, but it can be easily performed by Monte Carlo techniques.

Minimal solution of uncertainty propagation

The most general problem:

$$f(x_1, x_2, \ldots, x_n) \xrightarrow[Y_j = Y_j(X_1, X_2, \ldots, X_n)]{} f(y_1, y_2, \ldots, y_m)\,.$$

The 'minimal' solution: linear combinations, i.e.

$$\left\{ \begin{array}{l} \mathrm{E}(X_i) \\ \sigma(X_i) \\ \rho(X_i, X_{i'}) \end{array} \right. \xrightarrow[Y_j = c_{j0} + c_{j1}X_1 + c_{j2}X_2 + \cdots + c_{jn}X_n]{} \left\{ \begin{array}{l} \mathrm{E}(Y_j) \\ \sigma(Y_j) \\ \rho(Y_j, Y_{j'}) \end{array} \right.$$

But not forgetting the correlations!

Linear combination of independent variables

Simple, but instructive and important case:

- One (*output*) variable (Y) depending from many (*input*) quantities X_i, $i = 1, 2, \ldots, n$.

$$\begin{aligned} Y &= c_0 + c_1 X_1 + c_2 X_2 + \cdots + c_n X_n] \\ &= c_0 + \sum_i c_i\, X_i \end{aligned}$$

Linear combination of independent variables

Simple, but instructive and important case:

- One (*output*) variable (Y) depending from many (*input*) quantities X_i, $i = 1, 2, \ldots, n$.

$$\begin{aligned} Y &= c_0 + c_1 X_1 + c_2 X_2 + \cdots + c_n X_n] \\ &= c_0 + \sum_i c_i\, X_i \end{aligned}$$

Very general, well know property

$$\begin{aligned} \mathrm{E}[Y] &= c_0 + c_1 \mathrm{E}[X_1] + c_2 \mathrm{E}[X_2] + \cdots + c_n \mathrm{E}[X_n] \\ &= c_0 + \sum_i c_i \mathrm{E}[X_i] \end{aligned}$$

("the expected value is a linear operator")

Linear combination of independent variables

Simple, but instructive and important case:

- One (*output*) variable (Y) depending from many (*input*) quantities X_i, $i = 1, 2, \ldots, n$.

$$\begin{aligned} Y &= c_0 + c_1 X_1 + c_2 X_2 + \cdots + c_n X_n] \\ &= c_0 + \sum_i c_i\, X_i \end{aligned}$$

Very general, well know property

$$\begin{aligned} \mathrm{E}[Y] &= c_0 + c_1 \mathrm{E}[X_1] + c_2 \mathrm{E}[X_2] + \cdots + c_n \mathrm{E}[X_n] \\ &= c_0 + \sum_i c_i \mathrm{E}[X_i] \end{aligned}$$

("the expected value is a linear operator")

Less general (it holds only if X_i are independent) property:

$$\begin{aligned} \mathrm{Var}[Y] &= c_1^2 \mathrm{Var}[X_1] + c_2^2 \mathrm{Var}[X_2] + \cdots + c_n^2 \mathrm{E}[X_n] \\ &= \sum_i c_i^2 \mathrm{Var}[X_i] \end{aligned}$$

Linear combination of independent variables

These two properties (and the extension of the second in the case of *correlated input variable*) are the main reason to prefer, as mostly representative summaries of distributions,

- expected value
- standard deviation ($= \sqrt{\text{Var}}$)

having the same physical dimensions of the variable itself.

Linear combination of independent variables

These two properties (and the extension of the second in the case of *correlated input variable*) are the main reason to prefer, as mostly representative summaries of distributions,

- expected value
- standard deviation ($=\sqrt{\mathrm{Var}}$)

having the same physical dimensions of the variable itself.

The two numbers to put in the famous bottle...

Linear combination of independent variables

These two properties (and the extension of the second in the case of *correlated input variable*) are the main reason to prefer, as mostly representative summaries of distributions,

- expected value
- standard deviation ($=\sqrt{\text{Var}}$)

having the same physical dimensions of the variable itself.

The two numbers to put in the famous bottle...

Linear combination of independent variables

These two properties (and the extension of the second in the case of *correlated input variable*) are the main reason to prefer, as mostly representative summaries of distributions,

- expected value
- standard deviation ($=\sqrt{\text{Var}}$)

having the same physical dimensions of the variable itself.

The two numbers to put in the famous bottle...

(not by chance we use in QM *expectations* and *variances*)

Linear combination of independent variables

These two properties (and the extension of the second in the case of *correlated input variable*) are the main reason to prefer, as mostly representative summaries of distributions,

- expected value
- standard deviation ($=\sqrt{\mathrm{Var}}$)

having the same physical dimensions of the variable itself.

The two numbers to put in the famous bottle...

(not by chance we use in QM *expectations* and *variances*)

But what is the pdf of Y, $f(y)$?

Linear combination of independent variables

These two properties (and the extension of the second in the case of *correlated input variable*) are the main reason to prefer, as mostly representative summaries of distributions,

- expected value
- standard deviation ($=\sqrt{\text{Var}}$)

having the same physical dimensions of the variable itself.

The two numbers to put in the famous bottle...

(not by chance we use in QM *expectations* and *variances*)

But what is the pdf of Y, $f(y)$?
In general it is a difficult task (if we want a 'close form').

Linear combination of independent variables

These two properties (and the extension of the second in the case of *correlated input variable*) are the main reason to prefer, as mostly representative summaries of distributions,

- expected value
- standard deviation ($=\sqrt{\text{Var}}$)

having the same physical dimensions of the variable itself.

The two numbers to put in the famous bottle...

(not by chance we use in QM *expectations* and *variances*)

But what is the pdf of Y, $f(y)$?
In general it is a difficult task (if we want a 'close form').
Monte Carlo helps!

Linear combination of independent variables

These two properties (and the extension of the second in the case of *correlated input variable*) are the main reason to prefer, as mostly representative summaries of distributions,

- expected value
- standard deviation ($=\sqrt{\text{Var}}$)

having the same physical dimensions of the variable itself.

The two numbers to put in the famous bottle...

(not by chance we use in QM *expectations* and *variances*)

But what is the pdf of Y, $f(y)$?
In general it is a difficult task (if we want a 'close form').
Monte Carlo helps! (kein gedanken Experiment!)

Linear combination of Gaussian variables

A great, extra simplification occurs when all X_i are described by normal distributions, also of different μ_i and σ_i:

Linear combination of Gaussian variables

A great, extra simplification occurs when all X_i are described by normal distributions, also of different μ_i and σ_i:

"A linear combinations of Gaussians
is still Gaussian"!

Central Limit Theorem

Given $Y = \sum_{i=1}^{n} c_i X_i$

Central Limit Theorem

Given $Y = \sum_{i=1}^{n} c_i X_i$

- $\mathrm{E}[Y] = \sum_{i=1} c_i \, \mathrm{E}[X_i]$ is a very general property.

Central Limit Theorem

Given $Y = \sum_{i=1}^{n} c_i X_i$

- $\mathrm{E}[Y] = \sum_{i=1} c_i \, \mathrm{E}[X_i]$ is a very general property.
- $\sigma^2[Y] = \sum_{i=1} c_i^2 \sigma^2[X_i] = \sum_{i=1} c_i^2 \sigma_i^2$
  assumes independence of X_i.

Central Limit Theorem

Given $Y = \sum_{i=1}^{n} c_i X_i$

- $\text{E}[Y] = \sum_{i=1} c_i \, \text{E}[X_i]$ is a very general property.
- $\sigma^2[Y] = \sum_{i=1} c_i^2 \sigma^2[X_i] = \sum_{i=1} c_i^2 \sigma_i^2$
  assumes independence of X_i.

  But nothing yet about $f(y)$

Central Limit Theorem

Given $Y = \sum_{i=1}^{n} c_i X_i$

- $\mathrm{E}[Y] = \sum_{i=1} c_i\, \mathrm{E}[X_i]$ is a very general property.
- $\sigma^2[Y] = \sum_{i=1} c_i^2 \sigma^2[X_i] = \sum_{i=1} c_i^2 \sigma_i^2$
  assumes independence of X_i.

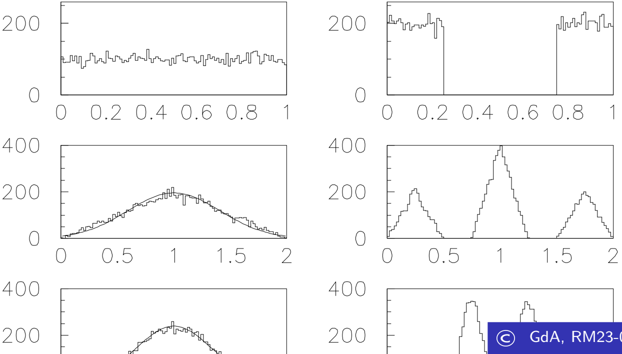
  But nothing yet about $f(y)$

Central Limit Theorem:

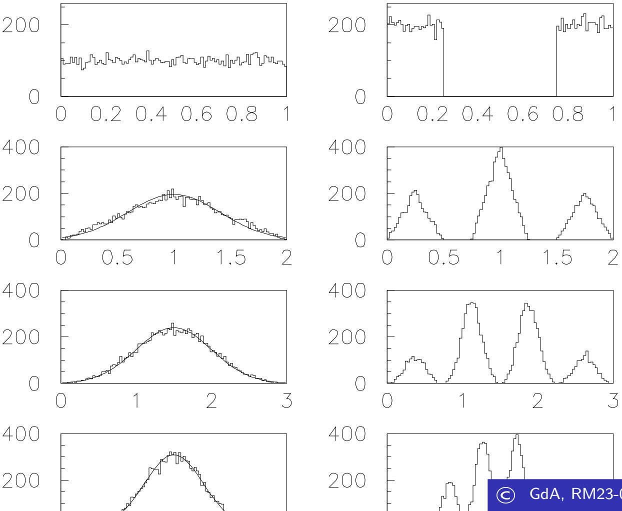
$$``n \rightarrow \infty" \Longrightarrow Y \sim \mathcal{N}\left(\sum_{i=1}^{n} c_i\, \mathrm{E}(X_i), \left(\sum_{i=1}^{n} c_i^2\, \sigma_i^2\right)^{\frac{1}{2}}\right)$$

if $c_i^2\, \sigma_i^2 << \sum_{i=1}^{n} c_i^2\, \sigma_i^2$ for all X_i not described by a Gaussian!
(i.e. a single non-Gaussian variable has not to dominate the uncertainty about Y.)

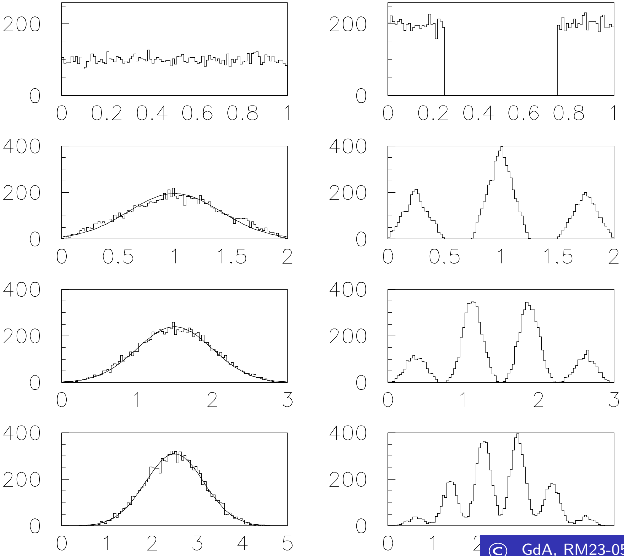
Central Limit Theorem: a cartoon ‘proof’



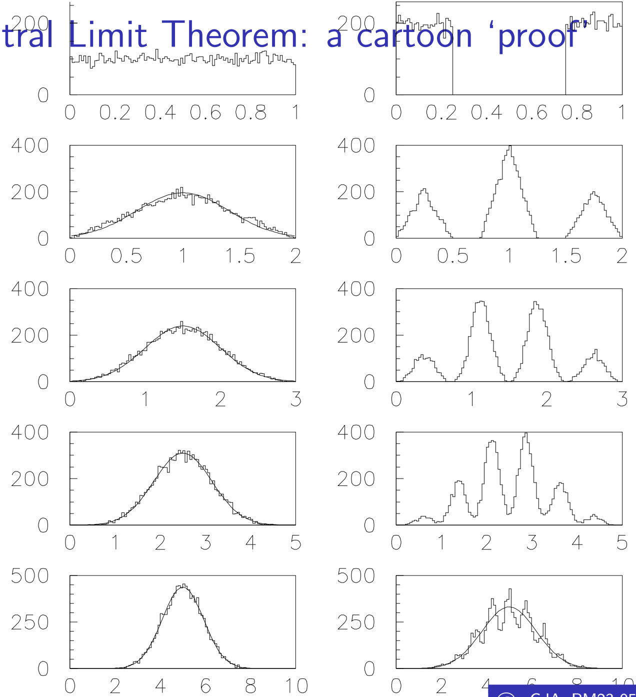
Central Limit Theorem: a cartoon ‘proof’



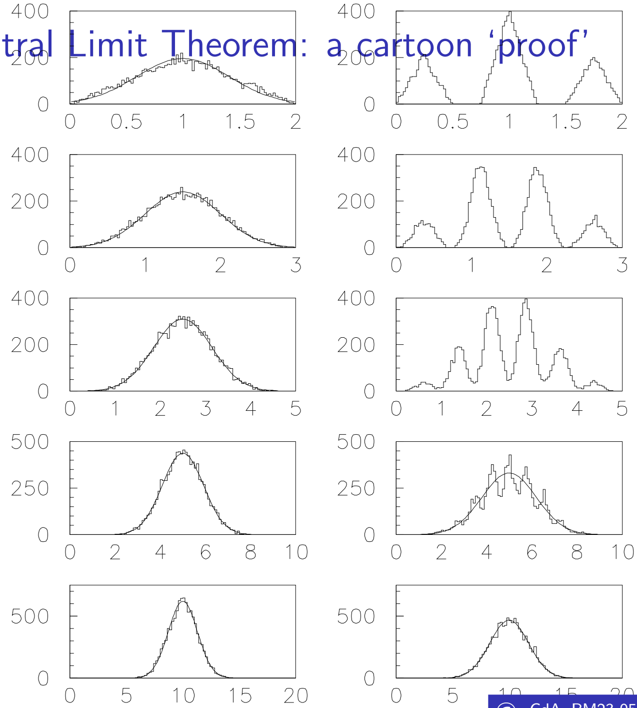
Central Limit Theorem: a cartoon ‘proof’



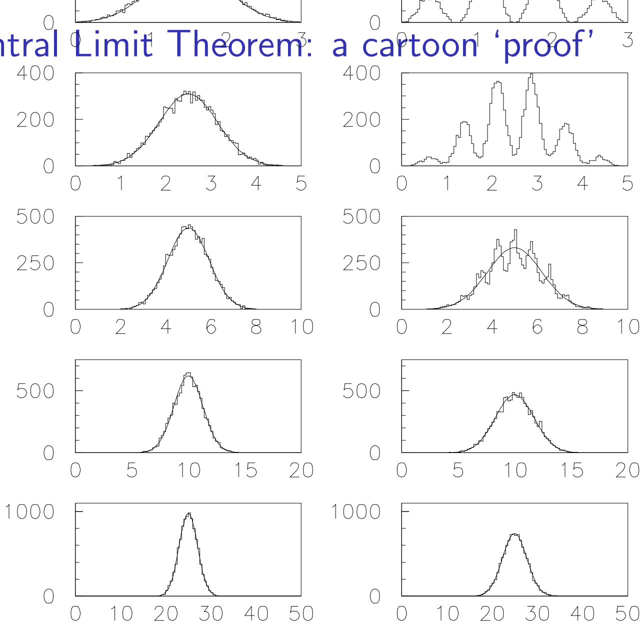
Central Limit Theorem: a cartoon 'proof'



Central Limit Theorem: a cartoon 'proof'

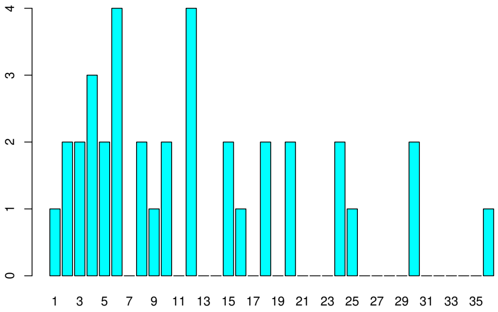


Central Limit Theorem: a cartoon 'proof'



Central Limit Theorem: another animated demonstration

Starting from the distribution of the product of the outcomes of two dice



(Unnormalized distribution)

⇒ R script

Sum of product of outcomes of *n* pairs of dice

```
pausa <- function() {
  cat ("\n >> press Enter to continue\n")
  scan()
}

outcomes <- as.vector(outer(1:6,1:6))
N <- 100
n <- 10000
sx <- rep(0, n)
for(ns in 1:N) {
  xi <- sample(outcomes, n, replace=TRUE)
  sx <- sx + xi
  hist(sx, nc=100, col='cyan', xlab='Sum X', freq=FALSE,
       main=sprintf("ns = %d; mean = %.1f, std = %.1f ",
                    ns, mean(sx), sd(sx)))
  pausa()
}
```