Progress with Uncertainty Quantification in Generic Monte Carlo Simulations

P. Saracco, M. Begalli and M.G. Pia
I.N.F.N. Genova (Italy)

NSS 2013 – Seoul

paolo.saracco@ge.infn.it
A simulation engine is needed when a given physical model is too much complicated to be analytically solved.
... but in a non idealized situation there are many sources of uncertainties.

Result of the simulation are affected by a complex mix of various sources of uncertainties.
... but in a non idealized situation there are many sources of uncertainties

Result of the simulation are affected by a complex mix of various sources of uncertainties

Neglect model uncertainties:
transport equation is the “true world” and there are no other approximations (like condensed path approximation) in its solution other than those induced by MC sampling

1st step to UQ

Physical data uncertainties

User’s data uncertainties

Solver (statistical) uncertainties

Result of the simulation are affected by a complex mix of various sources of uncertainties
... but in a non idealized situation there are many sources of uncertainties:

- MC users should have accurate info about confidence intervals (better PDFs) for each of the physical parameters used by the code (cross sections, models employed, ...)
- AND the possibility to make variations

Validation of physical data needed for simulation: MC users should have accurate info about confidence intervals (better PDFs) for each of the physical parameters used by the code (cross sections, models employed, ...)

2nd step to UQ

Because depending on the experimental configuration small variations in some parameter may result in large variations of the output or large variations in some other parameter may have no practical effect at all.
... but in a non idealized situation there are many sources of uncertainties.

Accurate study of the experimental configuration: MC users should select a proper restricted set of parameters (physical and/or user controlled) on which to perform an UQ attempt.

3rd step to UQ

Because as we shall see it is out of human possibilities (and probably meaningless) to perform UQ over possibly hundreds of different parameters.
Then (3rd step to UQ) we assume to have N independent source of uncertainty \( x_1, \ldots, x_N \) with their associate known PDFs (may be flat, normal or ...) and we want to derive the PDF for an observable \( Y \) in the range \((y, y+dy)\) from a MC simulation encompassing \( N_E \) events:

\[
G_{MC}(y) = \int_{-\infty}^{+\infty} d\bar{x} f_1(x_1) \cdots f_N(x_N) \exp \left[ -\frac{(y-y_0(\bar{x}))^2}{\sigma_{y_0}^2 / N_E} \right] \sqrt{\frac{N_E}{2\pi\sigma_{y_0}^2}}
\]

from the Central Limit Theorem.
This result trivially derives from probability composition

If the number of events is very large this distributions approaches

\[
G(y) = \int_{-\infty}^{+\infty} d\bar{x} f_1(x_1) \cdots f_N(x_N) \delta(y-y_0(\bar{x}))
\]

To do the job we need a way to easily handle this expression
Knowledge of $G(x)$ is the ultimate goal of UQ

Knowledge of $y_0(x_1,\ldots,x_N)$ required

USE MC to sample $y_0(x_1,\ldots,x_N)$

Reduced number of simulations needed

Caution: statistical errors un $y_0(x_1,\ldots,x_N)$

...with high statistics on $y_0(x_1,\ldots,x_N)$

Fixed values of $x_1,\ldots,x_N$
Parametric dependence of the physical output from input unknowns may be **nonlinear**

**Example (1D):** area of a circle \( y \) with (flat) uncertainty on its radius \( x \)
(here we exactly know \( y_0(x) = \pi x^2 \))

![Graph showing area as a function of radius with uncertainties.](image)

**METROPOLIS algorithm**

(a) Low MC statistics  
(b) Higher MC statistics

Thick black : exact result  
Histogram : sampled

For a single unknown we have no problems: it suffices to fit \( y_0(x) \) from few (say 10 or so) MC simulations


For many unknowns: **LINEARIZE** over different intervals
THE FINAL (not easy) TASK is then to handle something like:

\[ G(y) = \int_{-\infty}^{+\infty} d\vec{x} f_1(x_1) \cdots f_N(x_N) \delta(y - \bar{y}_0 - \sum_{k=1}^{N} a_k (x_k - \bar{x}_k)) \]

where parameters come from validation and/or problem specific analysis: \( \bar{x}_k \) are the means (or the modes) of the input unknowns, while

\[ \bar{y}_0 = (x_1, \ldots, x_N) \quad \text{and} \quad a_k = \frac{\partial y_0}{\partial x_k} \approx y_0(x_1, \ldots, x_k + \Delta, \ldots, x_N) - \bar{y}_0 \]

Statistical errors can be obtained with \( N+1 \) (high statistics) simulations (not thousands !!!)

UQ “reduces” to a well known (not solved) problem of classical probability theory: \( G(y) \) is the probability that the (weighted) sum of \( N \) independent arbitrary distributed stochastic variables assume some specific value.

In some cases solutions exist:
- All normal distributions
- All flat distributions
- All Lorentz distributions

NO GENERAL EXACT SOLUTION EXISTS
(common case: mixed flat and normal)
Find an approximate solution with prefixed accuracy

We found an exact expression holding for arbitrary polynomial distributions over finite different intervals

Any distribution – if necessary constrained to a finite interval with some known confidence – can be approximated with arbitrary prefixed accuracy by a polynomial (spline) one (Weierstrass approximation theorem)
TO SOLVE: the most efficient way to do this

To be done:
- Track statistical errors from parameter estimates (simulation)
- Track approximation errors for input PDFs
- Automate the process: a software user friendly tool

Approximation errors
Validation of physics ingredients

- The validation of the physics “ingredients” of Monte Carlo codes is a (complex, slow) still ongoing process
  - at least regarding Monte Carlo particle transport

- Quantitative estimates of input uncertainties (cross sections, angular distributions, BR etc.) are necessary for uncertainty propagation
  - A qualitative plot is not enough...

- Epistemic uncertainties are often embedded in the code, without being documented
Conclusion and outlook

We have established

- A novel conceptual approach
- A mathematical framework
- Calculation methods for single and many parameter uncertainties

These developments are applicable to Monte Carlo simulation in general Particle transport Event generators

Outlook

- Verification in a realistic experimental scenarios
- Application software system

In parallel, we pursue extensive Geant4 physics validation:
http://www.ge.infn.it/geant4/papers

Collaboration with HEP experiments is welcome!