Prerequisites for the Validation of Experiment and Theory

Lars Sonnenschein
CERN and LPNHE Paris, Universités Paris VI, VII

Abstract
In physics a better understanding of nature is achieved by a recursive interplay between experiment and theory. This requires a validation of both. On the theory side Monte-Carlo event generators can be validated by means of data from experiment. This data has to be corrected for detector effects to render an immediate comparison to event generators meaningful. A HepData database is available to retrieve published measurements including error correlation matrices from authors. Furthermore a validation framework Rivet is available in which authors are supposed to implement the necessary code to reproduce their published measurement exactly. To prevent any ambiguities this implementation should be accomplished at the time of publication. The constraints from published measurements are needed for further event generator development, of which experiments in turn will benefit in the next iteration.

1 Introduction

In high energy physics the ultimate goal of experiment and theory is a better understanding of nature. While the theory needs input from experiment for the verification or falsification of concurrent models the experiment needs input from theory for the prediction of observables, the understanding of scattering processes/production rates and the discrimination of instrumental effects and background processes from (new) physics. A recursive interplay takes place between experiment and theory where the experiment probes the description of nature provided by the theory, as schematically depicted in fig. 1. The intersection point where experiment and theory meet is the cross section. But before measurements can be compared to theory, the measurements have to be corrected for detector effects on the one hand and the models in which the theory is embedded have to be simulated on the other hand. To render the comparison between theory

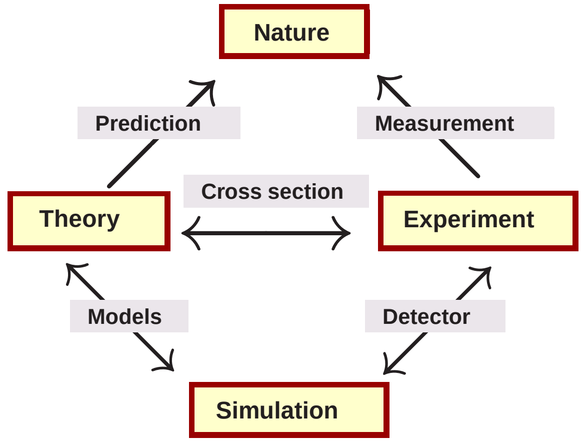

Fig. 1: Relations between theory, experiment, simulation and nature. The intersection point where experiment and theory meet is the cross section. While the theory makes predictions of nature and interfaces via models to the simulation the experiment measures nature and interfaces via the detector simulation or corrections obtained from data to the models.

and measurement meaningful the understanding (verification, validation and optimization) of Monte-Carlo event generation, simulation and experiment is crucial.

The need for the validation of experiment and theory is also documented by Sir Arthur Eddington's statement: "It is a good rule not to put overmuch confidence in a theory until it has been confirmed by observation. I hope I shall not shock the experimental physicists too much if I add that it is also a good rule not to put overmuch confidence in the observational results that are put forward *until they have been confirmed by theory*" (his italics).

2 Need for corrected data from experiment

The theory makes predictions to very few fixed orders (LO, NLO) plus resummation of radiation. More or less phenomenological models are needed for comparison with measurements. The models are implemented in Monte-Carlo event generators. They contain phenomenological parameters like e.g.:

- Parton shower termination parameters $p_\perp \min, m_{\min}$
- Lund string and cluster fragmentation parameters: string function parameters, mass
- Underlying event: primordial $k_\perp$, color reconnection parameters,
- Parton Distribution Functions (PDF's).

Therefore the models need to be validated and adjusted using real data from experiment. The data is coming form the HepData database [1] which is an archive of published HEP data from the last 30 years. It contains almost exclusively data which has been corrected for detector effects. Its focus is on cross section and similar measurements which makes the archive complementary to the Particle Data Group.

Authors who are publishing a measurement should remember to send their data to the HepData database. This data has to be corrected for detector effects (i.e. acceptance, efficiency and instrumental background) which corresponds to a correction to the hadronic final state or particle level. It is important that the data is not corrected any further to prevent the introduction of model dependencies since the models are supposed to be tested with the data among others. Only if corrected in this way the data can be always compared to Monte-Carlo event generators and it will be useful any time in the future. Otherwise the published measurement will be obsolete sooner or later (typically rather soon).

3 Reproducibility of published analyses

Before a comparison of the theory and models via simulation to data can be accomplished the published analyses have to be implemented and they have to match the publications exactly. Phenomenologists spend an enormous amount of time to reproduce published data analysis in all details, e.g. jet algorithm details and how the algorithm has been applied exactly. The publication might seem unambiguous at the time of writing. Experience shows, that this is no longer the case later on. The solution is the validation tool Rivet [2] which contains the analysis code and provides the real data for comparison. Rivet can be directly interfaced by means of the standardised event record format HepMC [3] to various Monte-Carlo event generators, e.g. via the interface package AGILe [4]. Authors of published corrected measurements (see last section for details on the correction) should implement their analysis into the Rivet framework and this at the time of publication to prevent any ambiguities. Only in this way an exact reproduction is guaranteed.

Present and past collider centre-of-mass energies provide unique points of operation. Event generator authors (of Herwig++, Pythia8, Sherpa, etc.) appreciate very much corrected analyses form the electron positron collider LEP where the hadronisation corrections turned out to be larger than the detector corrections. Important constraints on fragmentation models have been provided by LEP analyses. The most important ones have already been implemented into the Rivet validation framework.

Another important item to be mentioned within the context of reproducibility is the correlation between errors in the measurement. The matrices of correlated errors are typically only provided by analyses accomplished in the QCD group of experiments. This information has to be obtained on an event by event basis and can therefore not be recovered from published plots containing measured distributions. Thus it is extremely important to document this information, too.

Constraints from new published data corrected for detector effects are needed for further Monte-Carlo event generator development, the more the better. Experiments will benefit from it in the next iteration.

4 Summary

An important prerequisite for the validation of experiment and theory is that experiments correct their data for detector effects. In this way the data can be used at a later time point, when different or new models and/or Monte-Carlo event generators have to be validated and optimised. In the case of correlated errors it is also important that the experiment provides the covariance matrix, since this information can not be recovered from published plots containing measured distributions. Once a measurement is being published, the results should be send to the HepData database. The authors of the analysis should implement their analysis into the validation framework Rivet at the time of publication. In this way the usefulness of their measurement is guaranteed any time in the future. Experiments will benefit from the additional constraints imposed by their published analyses in the next iteration of event generator validation.

Acknowledgements

This work has been supported by the Marie Curie Program, in part under contract number MRTN-CT-2006-035606 and the HEPtools EU Marie Curie Research Training Network under contract number MRTN-CT-2006-035505.

References

[1] HepData database, `http://projects.hepforge.org/hepdata`.

[2] Validation framework Rivet, `http://projects.hepforge.org/rivet`.

[3] M. Dobbs, J. B. Hansen, L. Garren, and L. Sonnenschein, `http://lcgapp.cern.ch/project/simu/HepMC/20400/HepMC2_user_manual.pdf`.

[4] A Generator Interface Library, `http://projects.hepforge.org/agile`.