Statistics and Error and Data Analysis for Particle and Nuclear Physics

1 Introduction

Most students signing up for experimental physics courses envisage future work with complex devices, tools and electronics. On the other hand, students enlisting for theoretical physics studies typically think of days and nights behind the desk, working miracles with formulae. Very few realize that both experimental and theoretical achievements converge at one point: data analysis and hypothesis testing. In the fields of high energy and nuclear physics, most educational institutes do not host modern experimental facilities, thus data analysis is often the primary objective of doctoral theses for the experimentalists.

It has been observed that level of preparation of graduate students in the art of data analysis is unfortunately low. Often basic understanding of error treatment is missing, and knowledge of statistics is only preliminary, even though most students had some statistics courses during undergraduate studies. The reason is likely twofold: detachment of theory from actual practice, and underestimation of importance of these subjects in future research.

The goal of this course is to introduce the students to the specifics of data taking and further analysis in high energy and nuclear physics. Series of scenarios intend to demonstrate practical implications of principles of statistics, and to invite the students to do simple estimates themselves. Scenarios offer an introduction to key concepts of experiments and measurements, and an overview of basics of data and error analysis.

The course is worth 7.5 ECTS.

1.1 *Main Learning Objectives*

Main learning objectives of the course are:

- Learn specifics of a typical detector setup, how is it influenced by the physics requirements and what does it imply for data taking
- Learn basic procedures involved in measurement of physical observables
- Understand importance of proper data treatment in experiment
- Understand theory behind standard data treatment techniques
- Learn terminology used in relevant scientific literature and common practice
- Understand principles used in data processing and detector simulation software
- Learn how to work with relevant reference literature and online information sources

2 Scenarios

This compendium covers the following scenarios of the *Statistics and error analysis and data analysis for particle and nuclear physics (SED)* course:

- Introduction
- Principles of data correction
- Experimental errors
- Distributions
- Event reconstruction
- Understanding a sampling calorimeter

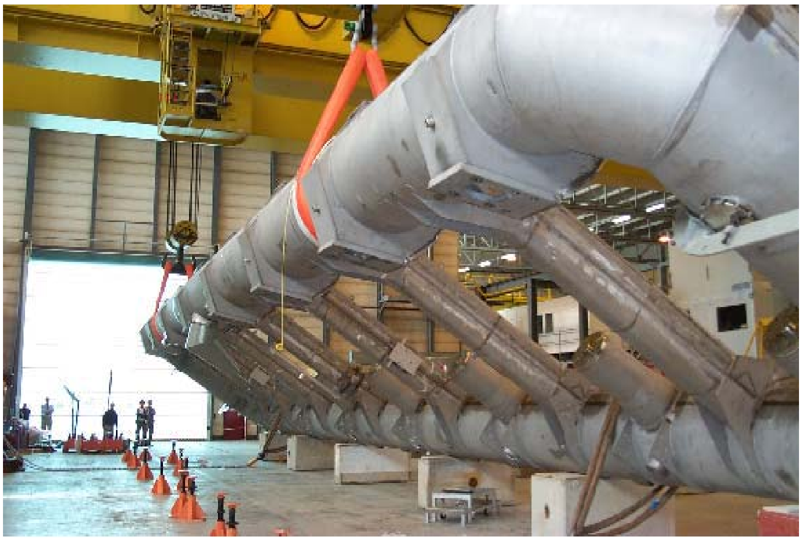
Each scenario is accompanied with the list of recommended literature and a text describing for supervisors the aims of the material.

Scenarios are organised in a sequential manner, starting with two introductory overviews, and continuing to more advanced scenarios concentrated on specific subjects. The first scenario bears specific marks of being one of the first PBL cycles offered to the group of students; hence it aims at engaging the students into active communication and into learning how to identify important questions for further studies. Following scenarios imply certain level of knowledge acquired by the students in due time. Every scenario introduces new terms that may be yet unknown to the students. Certain scenarios contain figures, for a better visual interpretation of the problem. Typically, scenarios do not foresee specific assignments except of the studies of the subject as such.

2.1 Introduction

2.1.1 Scenario: ATLAS Detector at LHC

On a rainy November afternoon, the security officer at CERN gates was surprised to see the expected delivery from Spain: a huge truck carrying 25-meter long, 5-meter wide construction of a round-cornered rectangle shape blocked the entire motorway, desperately trying to take turn and enter through the narrow gate. The students standing nearby were however not surprised: they started discussing enthusiastically that this must be the first of 8 *cryostats* that will hold the ATLAS *barrel toroid magnets*. This unique installation will provide a magnetic field of 4 Tesla, and will be one of the key components of a system that will facilitate efficient detection of decay products of awaited new particles, produced in collisions at the Large Hadron Collider.



2.1.2 Literature

1. ATLAS Detector and Physics Performance Technical Design Report, http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/TDR/access.html
2. The ATLAS Experiment Web Site, http://cern.ch/Atlas
3. C. Grupen, Particle Detectors (Cambridge University Press, 1996).

2.1.3 Aims of the material

The primarily goal of this scenario is to introduce the students to the specifics of a modern particle physics detector, through reasoning behind its construction and purpose.

Another goal is to make the students more familiar with the PBL approach, such that this scenario can be used as one of the first cycles for a group of novice students. It aims at stimulating lively discussions and yet keeping the students focused on identifying further learning goals.

It may be useful to distribute in mid-cycle hand-outs with drawings of the detector, or to provide a poster with a detailed layout. Visual appreciation of such a complex structure is very important, especially for the students who had no prior encounter with analogous installations.

Ability to be able to present and defend publicly their work is crucial in the very competitive field of High Energy Physics. Communication skills are also very important for the rest of the PBL course. Moreover, students at this early stage are not yet fluent in specific English terminology, thus the teacher must create an opportunity for them to learn and to use the terms. Most efficient approach is to assign every student to a public talk in front of the group.

As an assignment, each student is asked to prepare a 15-20 minutes slide presentation covering a subsystem of the ATLAS detector. Students select presentation subjects themselves, filling their names in a hand-out table as shown below:

Sub-system	Name
Pixel detector	
Semiconductor Tracker (SCT)	
Transition Radiation Tracker (TRT)	
Electromagnetic Calorimeter (LArg)	
Hadron Calorimeter (Tile)	
Monitored Drift Tubes (MDT)	
Cathode Strip Chambers (CSC)	
Resistive Plate Chambers (RPC)	
Thin Gap Chambers (TGC)	

2.2 Principles of data correction

2.2.1 Scenario: First encounter with real data

John is a student with a particle physics group. His first task is to analyse angular distribution of certain particles produced in a collision experiment and fit it with a function $\sim\cos^2\theta$. John receives experimental data samples that provide high statistics, and software tools that can extract information about registered events and particles. After applying a standard algorithm that selects suitable events and particles, John plots the angular distribution, only to discover that it does not look like a smooth $\sim\cos^2\theta$ function at all. Instead, he sees a quite irregular distribution, with distinct gaps and other odd shapes. John's supervisor suggests that the simplest solution to the problem is to apply a "*correction factor*", obtained with the help of "*Monte Carlo*". Unfortunately, available Monte Carlo samples do not contain sufficient statistics, and John is advised to request more of those through the collaboration's production system. The system offers to produce a rather confusing variety of samples, labelled "*evgen*", "*simul*", "*digit*", "*pileup*" and "*reco*". After some investigation, John manages to submit the requests in the correct order, and in a while receives a sufficient amount of Monte Carlo samples, that allow him to evaluate a satisfactory correction factor and finally recover the expected experimental $\sim\cos^2\theta$ distribution.

2.2.2 Literature

1. G. Cowan, Statistical Data Analysis (Oxford Science Publications 1998)
2. ATLAS Detector and Physics Performance Technical Design Report, http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/TDR/access.html

2.2.3 Aims of the material

The goal of this scenario is to introduce the students to the fact that detectors have large inefficiencies, which must be taken into account while analysing the data. The scenario focuses on a rather simple approach that relies on Monte Carlo for evaluating the correction factor. The $\cos^2\theta$ observable is chosen not to illustrate a specific process, but to indicate a smooth featureless distribution, suitable for this correction procedure. For more advanced students, more demanding observables and more appropriate unfolding methods may be mentioned.

Another purpose of this material is to make the students familiar with specific terminology used to refer to various types of Monte Carlo-produced data. Such terminology is not necessarily the same across the experiments, but certain similarities are always present. Moreover, this offers a way to learn about different components stages in Monte Carlo and real data treatment, such as an original particle generation, detector simulation and reconstruction.

Reference literature in this area is rather scarce, as no common methodology exists.