Data Analysis, advanced lectures by G. Macharashvili

High Energy Physics Institute of TSU

First part (14 Lectures, 2*14=28 hours):

I. Basics of probability theory.

1. Lecture (2 hours) Purpose and plan of lectures. Internet resources. Overview of software libraries. Literature review. Recommendations (root, gsl, scipy, ...). Definition of probability and conditional probability. Concept of random variable. Bayes definition.
2. Lecture (2 hours) Continuous and discrete random variables. Distribution function, properties and characteristics (parameters). Characteristic function, moments, functions of random variables.
3. Lecture (2 hours) Discrete distribution functions: binomial, multinomial, Poisson, geometric, and others. Continuous distribution functions: uniform, Gaussian (normal), exponential, χ^2, Student's, beta, Cauchy (Breit-Wigner), Landau, Maxwell and others. Connections between distributions and similar issues.
4. Lecture (2 hours) Transformations of random variables. Demonstration examples. Sums of random variables. Demonstration examples. Multidimensional random variables. Covariance and Correlation.

II. Central limit theorem. Statistical analysis methods.

1. Lecture (2 hours) Terminology, basic concepts: hypothesis tests, parameter estimates. Other methods of analysis. Population, sampling, empirical distribution function. Sampling characteristics and their evaluations. Likelihood function. Averaging of independent and correlated data.
2. Lecture (2 hours) Concept of statistical hypothesis. Hypothesis tests: Student's, Kolmogorov-Smirnov, Fisher's, Wilcoxon's, Pearson's (χ^2), test with p-value, and others. Two samples comparison test.
3. Lecture (2 hours) Point estimation of distribution function parameters. Maximum likelihood method. Parameter estimation method of least squares. Robust methods.
4. Lecture (2 hours) The distribution function parameters estimation with intervals. Asymmetric intervals. Parameter estimation with a one-sided limit. Example: searching for rare processes.
5. Lecture (2 hours) Histogram method. Histogram as a multinomial distribution. Estimation of sampling parameters with a histogram.

III. Uncertainty analysis.

1. Lecture (2 hours) Characteristics of measurements in particle physics. Terminology. Experiment, measurements, measurement errors. examples. Main characteristics of detectors. Performance, background, stability, resolution, dead-time and more.

2. Lecture (2 hours) Statistical errors. Variance and Standard Deviation. Systematic errors and their possible sources. Combining statistical and systematic errors.
3. Lecture (2 hours) Error propagation linear model. Demonstration example (fitting). A multidimensional case. Consideration of covariance. Error propagation by means of Monte Carlo method.
4. Lecture (2 hours) Characteristic uncertainties of elementary particle detectors. Overview of different experimental setups.

IV. (2 hours) *Overview of test for the exam. Overview of assignments. Consultation, discussion.*

Second part (8 Lectures, 2*8=16 hours):

V. Multivariate analysis.

1. Lecture (2 hours) Multivariate representation of data, definitions. Elements of Matrix Algebra: Definitions, Terminology, Diagonalization, Factorization, Eigenvalues and Eigenvectors.
2. Lecture (2 hours) Covariance matrix as a symmetric, positive definite matrix. Independent and correlated variables. Estimation of the sampling covariance matrix. Dynamic (synchronous) evaluation. Averaging independent and correlated data.
3. Lecture (2 hours) Diagonalization. Principal component analysis (PCA). Example. Feedforward neural networks. Examples (classification, function approximation).
4. Lecture (2 hours) Genetic algorithm. Fitting demo example. Kernel method. Examples. blind analysis.

VI. Monte Carlo method.

1. Lecture (2 hours) Model concept. Simulation, general overview. Random variable generators, general overview. Histogram method. geant4 work environment overview.
2. Lecture (2 hours) Models of physical processes. Transportation of charged particles in the medium. Propagation of gamma quanta in matter.
3. Lecture (2 hours) Simulation with a physical model. Demonstration of the general principle of hadronic therapy and the PET scanner. geant4 demo code.

VII. (2 hours) *Overview of test for the exam. Overview of assignments. Consultation, discussion.*

Literature

1. Textbook: G. Macharashvili. Data Analysis. Tbilisi 2022.