Advanced Methods for Modelling Physical Phenomena: from data gathering & analysis to hypothesis testing and parameter estimation

Sergey Vinogradov

Lebedev Institute and Moscow Physics Engineering Institute

Part 1. General approaches and considerations

Section 1. Experimental data and random variables: representation and statistics

1. Experimental and theoretical cycles and developments in physical studies
    a. Observation
    b. Experiment
    c. Model
2. Representation of experimental data and statistics
    a. Mean and standard deviation
    b. Empirical distribution
    c. Histogram
3. Probabilistic grounds for experimental data analysis
    a. Random variable
    b. Probability distribution
    c. Moments
4. Probability distribution functions
    a. PMF
    b. PDF
    c. CDF
5. Statistics of random variable
    a. Expectation and Variance
    b. Co-variance and Correlation
    c. Characteristic and generating functions
6. Stochastic process, random functions, waveforms and statistics
    a. From random variable to random function
    b. Averaging, ergodicity, stationarity
    c. Transient processes
    d. Nonlinear processes
7. Examples of stochastic processes
    a. Poisson
    b. Markov
    c. Cox

Section 2. Analytical modelling of stochastic processes and phenomena

8. Challenges, drawbacks and benefits of analytical modeling
    a. High risk – high reward area
    b. Analytics vs Monte Carlo
9. How to build a model of a sequence of stochastic processes
10. Distribution of sum of independent identically distributed random variables
    a. Moments of the sum
    b. PMF and PDF of the sum
    c. Random sum moments: Burgess variance theorem
    d. Random sum distribution: Laplace transform
11. Distribution of ordered random variables: order statistics approach
    a. Erlang distribution of arrival times of stationary Poisson process
    b. Order statistics of time-varying arrival times

Section 3. Monte Carlo modelling of stochastic processes and phenomena

12. Challenges, drawbacks and benefits of Monte Carlo modeling
13. Verification of MC modelling results

Section 4. Analysis of experimental and modelling results

14. Measurement errors
    a. Statistical
    b. Systematic
    c. Error propagation
15. Cramer-Rao lower bound on parameter estimation
    a. Fisher information integral
    b. Cramer-Rao lower bound
    c. Random number of events challenge
16. Empirical distribution function vs histogram: in-depth comparison
    a. Histogram binning challenge and its optimization
    b. Smirnov-Kolmogorov test of distance between distributions
    c. Benefits of empirical distribution function approach
17. Approaches to hypothesis testing and parameter estimation
    d. Least squares fitting
    e. Chi squared test
    f. Maximum Likelihood approach
    g. Maximum Likelihood and “goodness-of-fit”
    h. Smirnov-Kolmogorov test of “goodness-of-fit”

Part 2. Probabilistic modelling and analysis of photon-number-resolving detection and Silicon Photomultipliers

Section 1. Introduction to photon-number-resolving detection and SiPM

1. Physics, concepts, and designs of SiPM
2. Specific features of SiPM: correlated noise, nonlinearity

3. Measurement and characterization of photon-number-resolving detectors
4. Application areas and large projects with SiPM
5. State-of-art, challenges, and perspectives of SiPM technology

Section 2. Probabilistic modelling of SiPM response and performance in photon number resolution

6. How probability distributions of signal and noise are transformed from input to output
    a. Photo-conversion (binomial)
    b. Multiplication (excess noise)
    c. Dark counts (Poisson)
    d. Correlated events (crosstalk, afterpulsing)
7. Modelling of probability distributions of correlated events
    a. Geometric chain model
    b. Compound Poisson distribution
    c. Branching Poisson process model
    d. Generalized Poisson distribution
8. Analysis of experiments using Generalized Poisson distribution model of SiPM response
    a. Results of S. Vinogradov' studies
    b. Results of FBK and DESY studies
    c. Results of CERN CMS studies
    d. Application of the model to radiation-degraded SiPMs

Section 3. Probabilistic modelling of SiPM timing and performance in time resolution

9. How probability distribution of signal and noise event times are transformed by photon detection processes
    a. Poisson point process
    b. Super-Poisson (bunched) process affected by SiPM correlated events
    c. Sub-Poisson (anti-bunched) process affected by SiPM pixel recovery time
10. Analysis of interarrival time distribution for dark and correlated events
    a. Histogram approach
    b. Empirical Complimentary CDF approach
11. Single-photon time resolution measurement using multi-photon transit time histogram
12. Calibration of gain of noisy low gain photomultipliers using transit time histogram
13. Modelling of SiPM time resolution
    a. Filtered marked point process approach
    b. Scintillation pulse detection (TOF PET)
    c. Laser / LED pulse detection (LIDAR)
    d. Contributions of baseline fluctuations due to electronic noise and DCR

Section 4. Probabilistic modelling of SiPM nonlinearity and saturation

14. General consideration of nonlinearity as a source of random losses and noises
15. Binomial nonlinearity – detection of short light pulses
    a. Conventional model: balls (Poisson photons) in bins (pixels)
    b. Adjustments to crosstalk
    c. Adjustments to recovery
16. Recovery nonlinearity – detection of long light pulses

a. Conventional model: non-paralizible counting with dead time
b. Advanced model of exponential recovery process
c. Advanced model of Markov reward-renewal process
d. Recovery nonlinearity – detection of long light pulses

17. Excess noise factor of nonlinearity

Section 5. Analysis of applicability and competitiveness of photodetectors

18. Total ENF and Detective Quantum Efficiency as FOM for detector performance
    a. Noise-affected lower bound
    b. Nonlinearity-affected upper bound
    c. Optimal operating range
    d. Dynamic range and detective range
19. Analysis of performance of modern photodetectors and SiPMs
    a. PIN
    b. PMT
    c. APD
    d. SiPM
    e. SiPMs of various designs and vendors
20. Application analysis of SiPM competitiveness and potential
    a. TOF PET
    b. LIDAR

Big Data Analytics: tools & infrastructure

Marco Landoni
Istituto Nazionale di Astrofisica

Lezione 1 (4 ore in tutto, 2+2)
i) Introduzione ai BigData. Definizione ed esempi pratici sia in campo industriale che delle fisica
ii) Tecniche di gestione dei dati in regime BigData
iii) Maggiori framework conosciuti (Hadoop, Hive, Spark)
iv) Applicazione a un dataset reale in ambiente locale

Lezione 2 (4 ore in tutto, 2+2)
i) Approccio distribuito e cloud Computing. Accenno a gestione dei dati attraverso database non relazionali.
ii) Servizi offerti dalle piattaforme commerciali (Google ad Amazon) per i problemi di BIgData e di Machine Learning.
iii) Architetture e servizi cloud per data preparation e data cleaning.
iv) Esempi di applicazioni. Il caso di Google DataPrep

Lezione 3 (4 ore in tutto, 2+2)
i) Dalla data preparation al data mining e analytics. Il caso di Google BigQuery e Google Analytics. Applicazioni in ambito di Astrofisica delle alte energie.
ii) Calcolo Distribuito con Amazon. Il caso di Hubble Space Telescope. Analizzare un catalogo di una missione spaziale in 15 minuti.

Lezione 4 (4 ore in tutto)
i) Machine Learning. Accenni e prime proprietà
ii) Il problema della classificazione. Classificatori lineari Bayesiani e a massima verosimiglianza. Classificatori basati sulla distanza e su vettori di supporto (Support Vector Machine)
iii) Alberi di decisione e random forests. Adaboost.
iv) Introduzione al concetto di rete neurale
v) I tools messi a disposizione dalle piattaforme Cloud commericali per Machine-Learning-off the shelf.