Exam problems

Statistical Analysis of Experimental Data 2024/2025

Please note:

- each problem should be uploaded to Kampus platform as a separate PDF file (mandatory; notebook printout is OK). Please upload the notebook file itself when needed or give Google Colab link in a comment;
- solutions have to include description and justification of the approach used, as well as discussion of the presented results;
- solutions should be uploaded until **Sunday, February 9, 23:55 CET**;
- by uploading the solutions to Kampus you declare that they resulted from your own work and that you have not shared nor discussed them with anyone.

Problem 1

Consider a 3D cube with side a=1. "Deep core" inside the cube is defined as an intersection of eight spheres with radius R=1 centered at the cube vertices (one can also define it as a set of points inside the cube which are closer than R to all cubes vertices). Using Monte Carlo method, find the volume of the core. What is the number of MC generations required to get the value of the volume with 10^{-2} relative precision?

Problem 2

Assume you have collection of new one euro coins from country A and from country B, with masses m_A and m_B, respectively (we assume the coin mass is fixed for each country). You use electronic scale to measure the following four mass combinations:

$$m_1 = m_A + m_B \qquad m_2 = m_A + 2\,m_B$$
$$m_3 = 2\,m_A + m_B \qquad m_4 = 2\,m_A + 2\,m_B$$

The uncertainty of single mass measurement is σ. Assuming measurements are independent:

- find the optimal estimates of m_A and m_B, based on the maximum-likelihood method;
- find the corresponding uncertainties;
- how would your results change, if you assume the two masses are the same, $m_A = m_B$?

Problem 3

Consider search for new exotic scalar production in electron-positron scattering. Scalar S can be produced (together with the Z boson) in the scalar-strahlung process, $e^+e^- \to Z\,S$ while the main background comes from the Z boson pair production, $e^+e^- \to Z\,Z$. Events of S production can be selected based on the mass measurement, which is done with two independent methods: from direct reconstruction with uncertainty $\sigma_{dir} = 5\,\text{GeV}$ and using the so called recoil mass method (from second Z decay to leptons) with uncertainty $\sigma_{rec} = 3\,\text{GeV}$. Assuming the scalar mass is $M_S = 95$ GeV and the mass of the Z boson is $M_Z = 91$ GeV:

- define variable (give analytical formula) which should be used for optimal separation of signal events (S production) from Z production background. What is the interpretation of this variable?
- by changing the cut imposed on this variable, find the cut value resulting in the highest signal significance.

Signal significance (in the small signal approximation) is given by the ratio of the signal event number to the square root of the background estimate, $\frac{S}{\sqrt{B}}$, after selection cut. Assume normal distributions for the mass measurements.

Problem 4

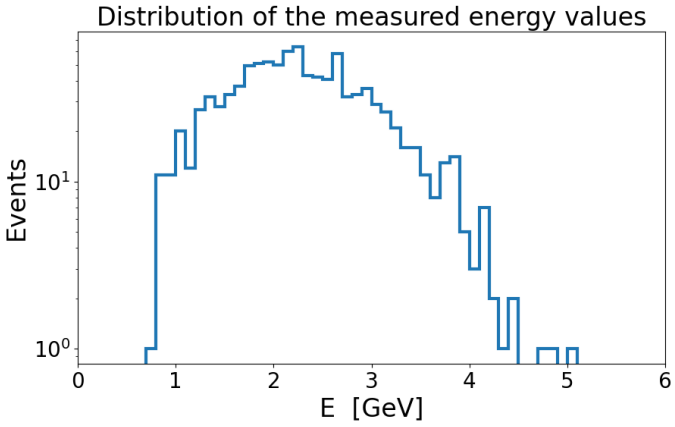
Calorimeter response to particle of given energy E_p (in GeV) is described by the Gamma distribution with mean and variance given by:

$$\langle E \rangle = E_p \qquad \text{and} \qquad \sigma_E^2 = 0.12 \text{ GeV} \cdot E_p$$

Calorimeter is put in the particle beam with flat energy spectra between E_1 and E_2:

$$p(E_p) = \begin{cases} \frac{1}{E_2 - E_1} & \text{for} \quad E_1 < E_p < E_2 \\ 0 & \text{otherwise} \end{cases}$$

Based on the 1000 measurements collected (as shown on the figure below; file available for download) extract beam energy spectra parameters (E_1 and E_2) and estimate their uncertainties.



Problem 5

Consider event selection in N-dimensional space ($\mathcal{R}^N$). Background events are uniformly distributed in a cube with side of $2\,a$, centered in the centre of the coordinate frame. Signal events are uniformly distributed inside a sphere of radius R, also centered in the centre of the coordinate frame. We assume $R \leq a$, i.e. sphere is inside the cube.

- what is the optimal approach to select signal enreached sample in this problem?
- calculate (the analytical formula for) the ROC function for this problem. What is the best value for the selection cut?
- generate required data sets and train one of the sklearn (or other library) machine-learning algorithms to discriminate between signal and background events for $N = 4$, $a = 1$ and $R = 0.8$. Compare the obtaioned ROC curve with your analytical result. Comment on the possible discrepancies.