Time	Group	Submission in Moodle; Mails with subject: [SMD2022]
Th.12:15–13:00	A	lukas.beiske@udo.edu and jean-marco.alameddine@udo.edu
Fr. 8:15–9:00	B	samuel.haefs@udo.edu and stefan.froese@udo.edu
Fr. 10:15–11:00	C	david.venker@udo.edu and lucas.witthaus@udo.edu

Exercise 13 *Kolmogorov–Smirnov-Test* **5 p.**

In this task, you investigate the similarity of the Poisson and Gaussian distributions using the Kolmogorov–Smirnov test.

(a) What values do you have to choose for μ and σ of a Gaussian distribution so that it is as similar as possible to a Poisson distribution with expected value λ?

(b) Implement the two-sample Kolmogorov–Smirnov test for binned data.

(c) The two-sample Kolmogorov–Smirnov test checks the null hypothesis H_0, whether the two samples stem from the same probability distribution. Investigate at which expected value λ the Poisson and Gaussian distributions are so similar that the Kolmogorov–Smirnov test can no longer distinguish between the two. To do this, draw 10 000 random numbers each from a Poisson distribution and from the corresponding Gaussian distribution for a λ to be tested. Consider the following:

- Round the values drawn from the Gaussian distribution to whole numbers.
- Use 100 bins each in the interval $[\mu - 5\sigma, \mu + 5\sigma]$.
- Determine by iteration the value for λ from which you can no longer reject H_0 on the basis of the Kolmogorov–Smirnov test at a confidence level of $\alpha = 5\,\%$.

(d) Determine the value for λ for the confidence levels $\alpha = 2.5\,\%$ and $\alpha = 0.1\,\%$ analogously.

Exercise 14 *Balloon Experiment* **5 p.**

In an experiment to measure the flux of cosmic rays in the upper atmosphere, protons with an energy between 1 GeV and 100 GeV are counted over a period of one hour from a flying balloon. Over a period of one week, a measurement run of one hour duration is made every day. The measured data are:

Day	1	2	3	4	5	6	7
Counts	4135	4202	4203	4218	4227	4231	4310

(a) Assume that the cosmic ray flux is constant over the measurement period. Calculate the most probable count rate using the maximum likelihood method.

(b) Your colleague looks at the readings and hypothesizes that the cosmic ray flux is experiencing a dramatic increase. Assume a linearly increasing flux and calculate numerically the most probable flux parameters using the maximum likelihood method.

(c) Calculate the significance of his observation using a likelihood ratio test. Evaluate the significance achieved. *Hint:* Assume that Wilks' theorem is valid here. Why can you assume this?

(d) Your colleague performs another measurement a week later to support his thesis. His measurement results in

Day	14
Counts	4402

Calculate (a) to (c) again for this new data set.

(e) What is the methodological problem with exercise d)'s approach? Why should you not publish these results, even if the significance is higher then some preset threshold (e.g. 3 or 5σ)?