Exercises - Introduction to Probability and Statistics

(1) For the following probability density

$$P(x) = xe^{-x} \qquad 0 \le x < \infty$$

(a) Find the mean and standard deviation of x. What is the probability content in the interval (mean-standard deviation, mean+standard deviation).
(b) Find the median and 68 % central interval
(c) Find the mode and 68 % smallest interval

(2) Consider the data in the table: Starting with a flat prior for each energy, find an

Energy	Trials	Successes
0.5	100	0
1.0	100	4
1.5	100	20
2.0	100	58
2.5	100	92
3.0	1000	987
3.5	1000	995
4.0	1000	998

estimate for the efficiency (success parameter p) as well as an uncertainty. For the estimate of the parameter, take the mode of the posterior probability for p and use the smallest interval to find the 68 % probability range. Make a plot of the result.

(3) Analyze the data in the table from a frequentist perspective by finding the 90 % confidence level interval for p as a function of energy. Use the Central Interval to find the 90 % CL interval for p.

(4) 9 events are observed in an experiment modeled with a Poisson probability distribution.
(a) What is the 95 % probability lower limit on the Poisson expectation value ν ? Take a flat prior for your calculations.
(b) What is the 68 % confidence level interval for ν using the Neyman construction and the smallest interval definition?

(5) Repeat the previous exercise, assuming you had a known background of 3.2 events.
   (a) Find the Feldman-Cousins 68 % Confidence Level interval
   (b) Find the Neyman 68 % Confidence Level interval
   (c) Find the 68 % Credible interval for ν

(6) (Only required for the TUM students) In this problem, we look at the relationship between an unbinned likelihood and a binned Poisson probability. We start with a one dimensional density $f(x|\lambda)$ depending on a parameter λ and defined and normalized in a range $[a, b]$. n events are measured with x values $x_i \quad i = 1, , n$. The unbinned likelihood is defined as the product of the densities

$$\mathcal{L}(\lambda) = \prod_{i=1}^{n} f(x_i|\lambda) \ .$$

Now we consider that the interval $[a, b]$ is divided into K subintervals (bins). Take for the expectation in bin j

$$\nu_j = \int_{\Delta_j} f(x|\lambda) dx$$

where the integral is over the x range in interval j, which is denoted as Δ_j. Define the probability of the data as the product of the Poisson probabilities in each bin.

We consider the limit $K \to \infty$ and, if no two measurements have exactly the same value of x, then each bin will have either $n_j = 0$ or $n_j = 1$ event. Show that this leads to

$$\lim_{K\to\infty} \prod_{j=1}^{K} \frac{e^{-\nu_j} \nu_j^{n_j}}{n_j!} \propto \prod_{i=1}^{n} f(x_i|\lambda) \ .$$

I.e., the unbinned likelihood is proportional to the limit of the product of Poisson probabilities for an infinitely fine binning.

(7) (Only required for the TUM students) In the *Extended Likelihood* approach, the overall normalization is taken into account so as to have a penalty if the observed number of events is not in line with expectations. Referring to the notation in the previous problem, we define

$$\Lambda = \int_a^b f(x|\lambda) dx$$

The normalized unbinned likelihood is now

$$\mathcal{L}(\lambda)_{\text{normalized}} = \prod_{i=1}^{n} \frac{f(x_i|\lambda)}{\Lambda}$$

and the extended likelihood is defined as

$$\mathcal{L}(\lambda)_{\text{extended}} = \frac{\Lambda^n e^{-\Lambda}}{n!} \prod_{i=1}^{n} \frac{f(x_i|\lambda)}{\Lambda} .$$

Compare the value of the unbinned likelihood, the binned Poisson probability, and the binned Poisson probability for the data set in the table for the model $f(x) = 10x \quad 0 \leq x \leq 1$. Discuss the results.

Set	x_1	x_2	x_3	x_4
1	0.5	0.5	0.5	0.5
2	0.2	0.5	0.75	0.9

(8) In this exercise, we want to fit a Sigmoid function to the following data:

Energy (E_i)	Trials (N_i)	Successes (r_i)
0.5	100	0
1.0	100	4
1.5	100	22
2.0	100	55
2.5	100	80
3.0	100	97
3.5	100	99
4.0	100	99

The Sigmoid function to fit is defined as:

$$s(E|A, E_0) = \frac{1}{1 + e^{-A(E-E_0)}}$$

(a) Find the posterior probability distribution for the parameters (A, E_0). You should clearly state what priors you assumed and your reasoning for the choice.

(b) Define a suitable test statistic and find the frequentist 68 % Confidence Level region for (A, E_0).

(c) Repeat the analysis of the data with the function

$$\epsilon(E) = \sin(A(E - E_0))$$

Find the posterior probability distribution for the parameters (A, E_0) and the 68 % CL region for (A, E_0)

(d) discuss the results