Prof. Dr. Lukas Heinrich
Dr. Jakob Knollmüller

Homework Exercises

Exercise 2 - 1

You are meeting an old friend. As you are catching up, he tells you that he now has two children. What is the probability that these are two brothers?

a) If you have **no further information**?

b) If he tells you that **the first born is a boy**?

c) If he tells you that **at least one of them is a boy**?

d) If he tells you that **one of them is a boy born on a Monday**?

e) As in (d), **but a week has n days**?

f) Examine the cases $n = 1$ and $n = \infty$ and explain why some of the probabilities from (a)-(c) are reproduced! What happens in between these extremes?

Exercise 2 - 2

Bob flips a (probably) manipulated coin as long as he gets tails. The moment the coin lands with head up, he stops the tossing. n denotes the number of tails he got. The coins' probability to land with tail up may be μ. Bob's strategy is denoted by B.

a) Calculate $\mathcal{P}(n \mid \mu, B)$.

b) Calculate the expected number of tails, i.e. $E[n]_{\mathcal{P}(n|\mu,B)}$.

c) Bob performs one tossing experiment from a) and gets n tails in a row, which he tells Alice. So far Alice does not know how the experiment was conducted and likes to infer the unfairness of the coin.

- Until now Alice believes that Bob performed a coin toss experiment of predetermined length $n + 1$. This strategy is called A. Calculate the most probable μ using $\mathcal{P}(\mu \mid n, A)$.
- Now Bob tells Alice that he ended the tossing when he got the first head. She therefore infers the most probable μ using $\mathcal{P}(n \mid \mu, B)$ and Bayes Theorem. Calculate $\mathcal{P}(\mu \mid n, B)$.
- Compare the results and discuss wether the finding is surprising.

d) As Alice knows that the maximum of a non-symmetric probability distribution is not equal to its expectation value she uses a computer algebra system of her choice to plot the probability distribution of μ.
Calculate the posterior mean $E[\mu]_{\mathcal{P}(\mu|n,B)}$ and compare with your results from b).

Exercise 2 - 3

a) Consider some experiment in which we record events that originate from two distinct sources (x and y), each following independently a Poisson distribution with rates λ_x and λ_y, respectively. The detector observes their sum $z = x + y$. Show that this random variable also follows a Poisson distribution and express it in terms of λ_x and λ_y.

b) Next we want to consider a *thinned Poisson process*. Here we have a random number of occurrences, N, distributed according to a Poisson distribution with rate μ. Each of the N occurrences, X_n, can take on values of 1, with probability μ, or 0, with probability $(1 - \mu)$, following a Bernoulli distribution. We want to derive the probability distribution for

$$X = \sum_{n=1}^{N} X_n. \tag{1}$$

Show that the probability distribution is given by

$$\mathcal{P}(X) = \frac{e^{-\lambda\mu}(\lambda\mu)^X}{X!}. \tag{2}$$

This scenario, for example, corresponds to experiments that produce some kind of physical event, but the detector only has a certain probability to record an event.

Exercise 2 - 4

You have a stick of length 1 and you break it into two pieces. The breaking point appears at a random location.

a) What is the average length of the smaller of the two pieces?

b) Calculate the average length ratio of the the smaller to the larger piece.

c) Calculate the average length ratio of the larger to the smaller piece.

d) You want to break another stick, but this time you use too much force and it breaks into three pieces. What is the mean length of the smallest piece?

Exercise 2 - 5

As a child, Nora always looked forward to getting the toy that was in the cereal box. She opened a new box every week, and looked for the toy. There were N different kinds of toys and they are distributed equally among the boxes.

a) Consider Nora already found m toys ($0 \leq m \leq N$), what is the probability to find a new toy in the next box?

b) Given the m toys, what is the expectation value of boxes to find the next new toy?

c) What is the expectation value for the number weeks until she had at least one of each kind of toy?

Exercise 2 - 6

The goal of this exercise is the implementation of the linear regression of a curve. The data consists of multiple pairs of (x, y) and we assume y to be a noisy measurement of some underlying function $y = f(x)+n$. We assume independent, additive, Gaussian noise $n \sim \mathcal{N}(n|0, N)$ with known variance N (e.g. $N = 0.1$, but could depend on the chosen function $f(x)$). We want to learn this function in terms of the polynomial basis $f(x) = \sum_{i=0}^{M} a_i x^i$ up to some order M in terms of the coefficients a. A priori we assume the coefficients to follow a Gaussian distribution $a \sim \mathcal{P}(a) = \mathcal{N}(a|0, A)$ with unit variance ($A = \mathbb{1}$).

a) We start by generating data according to the described process. Choose some (reasonably smooth) function $f(x)$ as ground-truth (e.g. $\sin(x), e^x$, some polynomial, combinations of those). Generate several sampling locations x for this function (either randomly or equidistant, maybe $10 - 30$ locations). Generate the corresponding y by applying the function to x and add a noise contribution n, drawn from the corresponding Gaussian.

Prof. Dr. Lukas Heinrich
Dr. Jakob Knollmüller

b) Plot the obtained data points together with the true function and indicate the error bars in y-direction in terms of the standard deviation.

c) Implement the prior covariance A, the noise covariance N, and the design matrix $R_{ij} = x_i^j$ as matrices. (exact definition, see script)

d) Implement the posterior covariance $D = \left(R^T N^{-1} R + A^{-1}\right)^{-1}$ and calculate $j = R^T N^{-1} y$

e) Use the posterior covariance to calculate the posterior mean of the coefficients $m = DR^T N^{-1} y$.

f) Plot the function corresponding to the mean coefficients m together with the data and ground-truth.

g) Draw several posterior samples from the Gaussian $a \sim \mathcal{N}(a|m, D)$ and add the corresponding functions to the previous plot (thin, transparent lines) to illustrate the uncertainty.

h) Now we want to determine which polynomial order M we should use. Calculate evidences for polynomial orders $M \in [0, \ldots, 10]$ and compare their values. What is the best polynomial order for your setup? Plot the evidence for the different M.

Hint: The evidence $P(y|M)$ is:

$$\mathcal{P}(y|M) = \frac{|2\pi D|^{\frac{1}{2}}}{|2\pi N|^{\frac{1}{2}} |2\pi A|^{\frac{1}{2}}} e^{-\frac{1}{2} y^T N^{-1} y + \frac{1}{2} j^T D j}, \text{ with} \tag{3}$$

$$D = \left(R^T N^{-1} R + A^{-1}\right)^{-1} \quad \text{and} \quad j = R^T N^{-1} y \tag{4}$$

i) How does this optimal order depend on the noise level? Repeat the same analysis for several noise covariances (and corresponding data) and plot your results.