

---

# Exercise Collection – Advanced Risk Min

---

## Contents

<b>Lecture exercises</b>	<b>1</b>
Exercise 1: Entropy . . . . .	1
Exercise 2: Kullback-Leibler Divergence . . . . .	1

---

## Lecture exercises

### Exercise 1: Entropy

A fair dice is rolled at the same time as a fair coin is tossed. Let  $A$  be the number on the upper surface of the dice and let  $B$  describe the outcome of the coin toss, where

$$B = \begin{cases} 1, & \text{head,} \\ 0, & \text{tail.} \end{cases}$$

Two random variables  $X$  and  $Y$  are given by  $X = A + B$  and  $Y = A - B$ , respectively.

- (a) Calculate the entropies  $H(X)$  and  $H(Y)$ , the conditional entropies  $H(Y|X)$  and  $H(X|Y)$ , the joint entropy  $H(X, Y)$  and the mutual information  $I(X; Y)$ .
- (b) Show that, for independent discrete random variables  $X$  and  $Y$ ,

$$I(X; X + Y) - I(Y; X + Y) = H(X) - H(Y)$$

### Exercise 2: Kullback-Leibler Divergence

- (a) You want to approximate the binomial distribution with  $n$  number of trials and probability  $p$  with a Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ . To find a suitable distribution you investigate the Kullback-Leibler divergence (KLD) in terms of the parameters  $\theta = (\mu, \sigma^2)^\top$ .
  - (i) Write down the KLD for the given setup.
  - (ii) Derive the gradients with respect to  $\theta$ .
  - (iii) Is there an analytic solution for the optimal parameter setting? If yes, derive the corresponding solution. If no, give a short reasoning.
  - (iv) Independent of the previous exercise, state a numerical procedure to minimize the KLD.
- (b) Sample points according to the true distribution and visualize the KLD for different parameter settings of the Gaussian distribution (including the optimal one if available).
- (c) Create a surface plot with axes  $n$  and  $p$  and colour value equal to the KLD for the optimal normal distribution.

- (d) Based on the previous result,
- (i) how can the behaviour for varying  $p$  be explained?
  - (ii) how can the behaviour for varying  $n$  be explained?