PD Dr. Torsten Enßlin
Margret Westerkamp
Johannes Harth-Kitzerow

Exercise sheet 4

Exercise 4 - 1

A sequence of n coin tosses is performed and stored in a data vector $d^{(n)} = (d_1, \dots d_n) \in \{0,1\}^n$. The coin produced a head (denoted by a 1 in the data vector) with constant, but unknown frequency $f = P(d_i = 1|f) \in [0,1]$.

a) How many bits of extra information on f are provided by the data vector $d^{(n)} = (1, \dots 1)$ of only ones?

Hint: The extra information contained in a probability distribution $p(x)$ compared to a probability distribution $q(x)$ (in bits) is given by $\int \mathrm{d}x\, p(x) \log_2 \left(\frac{p(x)}{q(x)}\right)$. Furthermore you may use the following integral formulas:
$\int_0^1 dx\, x^n (1-x)^m = \frac{n!\, m!}{(n+m+1)!}$ for $n, m \in \mathbb{N}$, $\int_0^1 dx\, x^n \ln x = -\frac{1}{(n+1)^2}$ (2 points)

b) After how many such sequential heads did one obtain 10 bits of information on f? An accuracy of 10% is sufficient.

Hint: If $n > 10$ you can use $\frac{n}{n+1} \approx 1$. Use $2^{1/\ln 2} = 2^{\ln e / \ln 2} = 2^{\log_2 e} = e \approx 2.7$. (1 point)

c) How many bits on the outcome of the next toss is provided by a sequence of n heads? Provide also the asymptotic for $n \to \infty$!

Hint: It is helpful to guess the maximal amount of obtainable information before the detailed calculation is done. (2 points)

Exercise 4 - 2

You are in the setting from exercise 4-1. Write a programm that generates a data vector $d^{(n)} = (d_1, \dots d_n) \in \{0,1\}^n$ by performing a series of virtual coin flips with heads-probability f.

a) Print the posterior mean and variance of f given the data vector d and a flat prior on f. (optional)

b) Print how much information the data vector d provides about f compared to the prior. (optional)

c) Print how much information you get about f in one coin flip compared to your knowledge before the coin flip. (optional)

d) Try to find a situation where your posterior variance increases after a coin flip. How much information did you get about f in that step? How much total information (with respect to the prior) do you have about f after this coin flip compared to before? (optional)

Exercise 4 - 3

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 on tail may be f. Bob's strategy is denoted by by B.

a) Calculate $P(n \mid f, B)$ (1 point).

b) Calculate the expected number of toins, i.e. $\langle n \rangle_{(n|f,B)}$(2 points).

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 f using $P\left(f \mid n, A\right)$ (2 points).
- Now Bob tells Alice that he ended the tossing when he got the first head. She therefore infers the most probable f using $P\left(n \mid f, B\right)$ and Bayes Theorem. Calculate $P\left(f \mid n, B\right)$.
- Compare the results and discuss if the finding is surprising (1 point).

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 f.
Calculate $\langle f\rangle_{(f|n,B)}$ and compare with your results from *b*) (2 points).

Exercise 4 - 4

Imagine you would like to store a probability distribution $P(x)$ over two events $x \in \{0,1\}$ on a computer with very limited memory and precision. Of course it is enough to store only $P(0)$ since $P(1)$ can then be calculated through normalization, but you still have to round the numbers to the computers precision. Let X be the set of all numbers the computer can represent. Furthermore, let

$$\begin{aligned} q_{\text{low}} &= \max\{q|q \in X \wedge q \leq P(0)\} \\ \text{and } q_{\text{high}} &= \min\{q|q \in X \wedge q \geq P(0)\} \end{aligned}$$

i.e., q_{low} the highest number in X that is still lower than $P(0)$ and q_{high} the lowest number in X that is still higher than $P(0)$.

a) Derive a decision rule when it is better to round to q_{low} or q_{high} for general $P(0)$, q_{low} and q_{high} based on the rule that you want to loose the least amount of information of the original distribution $P(x)$. (2 points)

b) Using the decision rule of a), determine wether it is better to round

- $P(1) = 0.146$ to 0.1 or 0.2
- $P(1) = 0.01$ to 0 or 0.5?

(1 point)

This exercise sheet will be discussed during the exercises.
group A, Wednesday 16:00 - 18:00 Theresienstr. 39 - room B 101,
group B, Thursday, 10:00 - 12:00, Geschw.-Scholl-Pl. 1 (B) - room B 015
group C, Thursday, 16:00 - 18:00, Theresienstr. 37 - room A 449
backup slot, Wednesday, 8:00 - 10:00, Theresienstr. 37 - A 450
(to be used whenever Thursday is a vacation day)
www.mpa-garching.mpg.de/∼ensslin/lectures