Exercise sheet 4

Exercise 4 - 1

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

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

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)

Exercise 4 - 2

Consider the following coin toss experiment:

- A large number n of coin tosses are performed and the results are stored in a data vector $d^{(n)} = (d_1, \ldots, d_n) \in \{0, 1\}^n$, where 0 and 1 represent the possible outcomes head and tail.
- Individual tosses are independent from each other.
- All tosses are done with the same coin with an unknown bias $f \in [0, 1]$; i.e., $\mathcal{P}(d_i|f) = f^{d_i}(1-f)^{1-d_i}$.

Assume that a fraction $\bar{f}$ out of the n coin tosses yielded head.

a) Derive the Gaussian approximation of the PDF $\mathcal{P}(f|d^{(n)})$ around its maximum. – You can use a saddle point approximation; i.e., identify the maximum, and taylor-expand the (negative) logarithm of $\mathcal{P}(f|d^{(n)})$ around it up to second order in order to identify the variance of the Gaussian (3 points).

b) Use this Gaussian approximation to derive an approximation for $\mathcal{P}(d_{n+1}|d^{(n)})$.
Hint: You can assume that the Gaussian distribution is narrow enough such that the integration boundaries $[0, 1]$ can be replaced by $(-\infty, \infty)$ (2 points).

c) Now calculate the exact posterior mean for $\bar{f}$ and the exact expression for $\mathcal{P}(d_{n+1}|d^{(n)})$ (2 points).

Note: $\int_0^1 \mathrm{d}x\, x^\alpha(1-x)^\beta = \Gamma(\alpha+1)\Gamma(\beta+1)/\Gamma(\alpha+\beta+2)$, where Γ is the Gamma function.

Exercise 4 - 3

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 - 4

You are in the setting from exercise 4-3. 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)

This exercise sheet will be discussed during the exercises.
Group 01, Wednesday 16:00 - 18:00, online,
Group 02, Thursday, 10:00 - 12:00, online,
Group 03, Thursday, 16:00 - 18:00, online,
Please hand in your solution by Monday, May 10th 2021, noon online:
https://moodle.lmu.de/course/view.php?id=14930