Probability and statistics.

Master in Cognitive Science. Academic year 2025-2026.
Example sheet 3.

Exercise 1:
On a farm there are four ducks, four geese and two hens. During the night, two wolves arrive. Each wolf catches one bird at random. Let us denote as:

- X = the number of ducks caught,
- Y = the number of geese caught.

1. Determine the distribution of the random variable X.
2. Determine the distribution of Y.
3. Determine the joint distribution of the random variables X and Y.
4. Determine the marginal distribution of X and Y from their joint distribution.
5. Compute the probability for the event that the number of ducks caught equals the number of geese caught.
6. Are the random variables X and Y dependent?

Exercise 2:
Ursula and Vera have agreed to meet for lunch exactly at noon $(12:00)$. However, Ursula arrives U minutes late, and Vera arrives V minutes late. We assume their arrival times are independent, and both are uniformly distributed over the interval $[0, 60]$.

1. Write down the density functions f_U and f_V , and the joint density function $f_{U,V}$. Start from the individual densities and then deduce the joint density.
2. Calculate the probability that Ursula arrives before $12:20$. Use the distribution of U only.
3. Find geometrically the probability that Ursula arrives before $12:15$ and Vera arrives between $12:30$ and $12:45$.
4. Calculate the probability in (3) again, without resorting to geometry, from the probabilities of the individual events $\{f_U < 15\}$ and $\{30 < f_V < 45\}$. Exploit the independence of U and V.
5. Find geometrically the probability that Ursula arrives at least 30 minutes after Vera. Try fixing a value of V, and think what must U then be for the event to occur. After you have found and drawn the region, calculate its area. Recall the area of a triangle.

Remark: the integral of a constant function over a region equals that constant times the area of the region.

Exercise 3:
Our objective is to locate an airplane that has disappeared from the radar system. The search region is divided into four quadrants, labeled $1, 2, 3$, and 4. Based on prior information, the probabilities that the airplane is located in each quadrant are $0.5, 0.3, 0.1$, and 0.1, respectively.
When a quadrant is searched, the probability of successfully finding the airplane—provided it is actually in that quadrant—is 0.3. This detection probability is independent of all previous search attempts. If the airplane is not in the searched quadrant, it cannot be found during that search.

1. The airplane location is modelled as a random variable
$$\Theta \in \{1, 2, 3, 4\}.$$
   Give its prior distribution.

2. The search starts with quadrant 1. Let
$$X = \begin{cases} 1, & \text{the airplane is found,} \\ 0, & \text{else.} \end{cases}$$
   Determine the likelihood of $X = 0$.

3. The airplane is not found on the first attempt. Determine the posterior distribution of the airplane location based on this information. Do you think we should search quadrant 1 again?

4. We decide to search quadrant 1 again. Define Y as the indicator variable for the second search result:
$$Y = \begin{cases} 1, & \text{the airplane is found,} \\ 0, & \text{else.} \end{cases}$$
   The airplane is still missing. Which quadrant should we search next?

Exercise 4:
Scientists study the effect of fructose on mouse weights. They are equipped with a weighing machine A. The weight measurements suffer from a small error due to the lack of precision of the machine.

1. Model the errors with the random variable Y_A that follows a standard normal distribution. In other words, for a mouse of exact weight m, the weighing machine returns a measure $m + Y_A$. Compute the probability that a mouse weight will be overestimated by $0.5g$.

2. The machine A is now broken, you use another one, says the weighing machine B. The error measure with this machine now modeled as a random variable Y_B that follows a Normal distribution $\mathcal{N}(0, 22)$. What is the probability of overestimating the weight of a mouse by 0.5g?

3. Scientists want to improve the precision of the weighting machine B. They propose to repeat the measures taken on each mouse. How many measurements they should take to make sure to have a precision at least as good as that of the machine A.

4. Using Rjags. You will find on a Ametice a dataset with the weight measures from 18 mouses under fructose diet and 15 mouses under no fructose diet. Using Rjags, compute the posterior distribution of mouse weights for each group. Compare the mean weights of both groups. You can choose a normal prior on the mean and a gamma prior on the precision. How can you make these priors non informative ?

Exercise 5:
In this exercise, we model the memory retention over time. The usual experimental design consists in giving people many items to remember on a list. Their ability to remember these items is tested after different time periods.
There are many (neuro-scientifically motivated models) for the memory retention over time, we consider here an exponential decay of the form $\theta_t = \exp(-\alpha t) + \beta$, $\theta_t \in (0, 1)$. α is the rate of decay of information, and β corresponds to a baseline level of remembering that is assumed to remain even after very long time periods.
Download on Ametice the data. It is a table which reported 4 subjects tested on 18 items at 10 time intervals. Each entry of the table gives the number of correct memory recalls for each subject at each time interval.