Probability and statistics.

Master in Cognitive Science. Academic year 2025-2026.
Example sheet 4.

Exercise 1:
Law of large numbers (LLN)
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.

Exercise 2:
Comparison of two estimators of a proportion.
Consider an n-sample $\mathcal{X} = (X_1, \ldots, X_n)$ of i.i.d r.v. with Bernoulli distribution $Be(\theta)$. For estimating θ you can use $\hat{\theta}_1 = \bar{X}$ or also $\hat{\theta}_2 = \frac{Y + \sqrt{n}/2}{n + \sqrt{n}}$, where $Y = \sum_{i=1}^{n} X_i$.

1. Compute the expectation, variance and mean-square error of these two estimators.

2. Interpret these results.

3. Build a computer experiment in R to check your previous findings.

Exercise 3:
Central limit theorem (CLT)
Consider a simple model of neuronal firing where each neuron in a large population emits a spike (1) or not (0) in a short time window. Let X_i denote the spike of neuron i:

$$X_i = \begin{cases} 1 & \text{if neuron } i \text{ fires in the interval,} \\ 0 & \text{otherwise,} \end{cases}$$

with $\mathrm{P}(X_i = 1) = p$ and $\mathrm{P}(X_i = 0) = 1 - p$. Assume the neurons fire independently.

1. Compute $\mathrm{E}[X_i]$ and $\mathrm{Var}(X_i)$.

2. Let $S_n = \sum_{i=1}^{n} X_i$ be the total number of spikes in the population. Express the standardized variable

$$Z_n = \frac{S_n - np}{\sqrt{np(1-p)}}$$

and explain what the Central Limit Theorem tells us about its distribution as $n \to \infty$.

3. Suppose $p = 0.1$ and $n = 100$. Approximate the probability that at least 15 neurons fire simultaneously:

$$\mathrm{P}(S_n \geq 15)$$

using the normal approximation given by the CLT.

4*. Discuss the relevance of this result in interpreting mean population activity in neural recordings. Why is the CLT useful when modeling aggregate signals such as local field potentials or fMRI BOLD responses?