# Computational Statistics 732A89 – Spring 2025
# Computer Lab 1

Martin Andrae, Bayu Brahmantio, Frank Miller
Department of Computer and Information Science (IDA), Linköpings University

January 22, 2025

This computer laboratory is part of the examination for the Computational Statistics course. Create a group report (which is directly presentable, if you are a presenting group), on the solutions to the lab as a PDF file. Be concise and do not include unnecessary printouts and figures produced by the software and not required in the assignments.
**All R code should be included as an appendix to your report.**
A typical lab report should contain 2-4 pages of text plus some figures plus an appendix with codes. In the report, refer to all consulted sources and disclose all collaborations.
The report should be handed in via LISAM (or alternatively in case of problems by email) by **23:59 January 29, 2025** at the latest. Notice that there is a deadline for corrections 23:59 08 April 2025 and a final deadline of 23:59 29 April 2025 after which no submissions or corrections will be considered, and you will have to redo the missing labs next year. The seminar for this lab will take place **February 12, 2025**.
The report has to be written in English.

## Question 1: Maximization of a function in one variable

Consider the function

$$g(x) = \frac{\log(x+1)}{x^{3/2}+1}.$$

a. Plot the function $g(x)$ in the interval $[0, 4]$. What is your guess for the maximum point?

b. Compute the derivative $g'(x)$ of $g(x)$; recall the quotient rule that for $g(x) = \frac{u(x)}{v(x)}$, the derivative is $g'(x) = \frac{u'(x)v(x)-u(x)v'(x)}{(v(x))^2}$. Plot $g'$ in $[0, 4]$, and add a horizontal reference line at 0 to the plot.

c. Write your own `R` function applying the bisection method to $g'$ to find a local maximum of $g$ for a user-selected starting interval.

d. Write your own `R` function applying the secant method to $g'$ to find a local maximum of $g$ for a user-selected pair of starting values.

e. Run the functions in c. and d. for different starting intervals/pairs of starting values and check when they converge to the true maximum and when not. Discuss why. Compare the two methods also in terms of number of iterations used and programming effort required.

f. When you just should program one of them: Would you use bisection or secant, here? In general, for another function $g(x)$ to be maximized: When would you switch and use the other algorithm?

# Question 2: Computer arithmetics (variance)

A known formula for estimating the variance based on a vector of $n$ observations is

$$\mathrm{Var}(\vec{x}) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right)$$

a. Write your own R function, `myvar`, to estimate the variance in this way.

b. Generate a vector $x = (x_1, \ldots, x_{10000})$ with 10000 random numbers with mean $10^8$ and variance 1.

c. For each subset $X_i = \{x_1, \ldots, x_i\}$, $i = 1, \ldots, 10000$ compute the difference $Y_i = \texttt{myvar}(X_i) - \texttt{var}(X_i)$, where $\texttt{var}(X_i)$ is the standard variance estimation function in R. Plot the dependence $Y_i$ on $i$. Draw conclusions from this plot. How well does your function work? Can you explain the behaviour?

d. How can you better implement a variance estimator? Find and implement a formula that will give the same results as `var()`.